Shun Hing Institute of Advanced Engineering
信興高等工程研究所

Research Highlights
2005 - 2007

香 港 中 文 大 學
The Chinese University of Hong Kong

Published in January 2008, this report contains information known as December 2007.

For further information, please visit our website: http://www.shiae.cuhk.edu.hk

Foreword

In late October 2004, Dr William Mong, founder of the Shun Hing Group of companies and a staunch friend and supporter of the Chinese University for over three decades, was on campus to open the magnificent building that bears his name, and to inaugurate the Shun Hing Institute of Advanced Engineering (SHIAE). The building provides state-of-the-art teaching and research facilities for the engineering disciplines, and the Institute spearheads advanced research, connects the University with other major centres of engineering studies in the world, enables the latest developments in technology to be effectively transferred to the industries, and generally advances the cause of engineering studies in the community at large.

Three years have elapsed since that happy occasion and, with the support of Dr William Mong's munificence, The Chinese University of Hong Kong has gone from strength to strength in exploring the frontiers of advanced engineering research. As Director of SHIAE, it is my very pleasant duty to present to the international academic community as well as the general public the efforts of my colleagues in the past three years, and their achievements in their respective areas of excellence, in the form of the present volume, which reports on the research projects undertaken by members of the Institute from 2005 to 2007. Our success during the initial triennium is our way of expressing our deep gratitude to Dr William Mong, and as we record the fruit of our labours in print, we deem it most appropriate for this issue of the SHIAE report to be dedicated to Dr Mong, and published as a festschrift in honour of his eightieth birthday.

As an entrepreneur who has built a business empire of global dimensions, Dr Mong has contributed significantly to the development of tertiary education in Hong Kong, and not the least to high-end engineering research at The Chinese University of Hong Kong, amongst his many and frequent philanthropic endeavours. We at the Shun Hing Institute of Advanced Engineering are honoured to enjoy his support, and privileged to benefit from his generosity. Confucius has observed in the Analects that "The benevolent is blessed with longevity (仁者壽)." This is an apt description of the personality of Dr Mong and the good work he has done, and on behalf of SHIAE and the Chinese University, may I again offer our best wishes to Dr William Mong, and look forward to his continued interest and support in our forays into the boundless firmament of learning and research.

Professor Pak Chung Ching
Pro-Vice-Chancellor
Director, Shun Hing Institute of Advanced Engineering
The Chinese University of Hong Kong


信興高等工程研究所
Shun Hing Institute of Advanced Engineering

Content

Introduction of SHIAE

The Shun Hing Institute of Advanced Engineering is a strategic initiative of the Chinese University of Hong Kong dedicated to excellence and innovation in the research and development of advanced engineering areas relevant to Hong Kong and the region. It is established through the generous funding support of the Shun Hing Education and Charity Fund, a splendid example of crossover partnership of the private sector in state-of-the-art research at a tertiary institution. It is the first and excellent of its kind.

The Shun Hing Education and Charity Fund is a dedication of Dr. William Mong Man Wai and Mr. David Mong in enhancing education opportunity for the younger generation. The Fund has already sponsored numerous education and research programmes in Hong Kong, the Mainland, and overseas education institutions. Himself an engineer and believing in advancing quality of life through developments of science and technology, Dr. Mong pledges his support to the establishment of this Institute almost instantaneously.

Centre of Excellence at CUHK

The Chinese University of Hong Kong is an internationally renowned institution of higher learning devoted to quality teaching and both academic and applied research. The University has established ten research institutes and a number of research centres with a view to pursuing up-front research endeavours with focused goals and objectives. The Shun Hing Institute of Advanced Engineering constitutes a crucial addition to the research infrastructure of the Chinese University which is committed to exciting research programmes in advanced engineering areas.

As a new centre of excellence at the Chinese University of Hong Kong, the Institute is to support greater regional and international research collaborations, and strives to attract talents from the world over to achieve greater internationalization, a vision strongly advocated by Professor Lawrence Lau, the Vice Chancellor.

Commitment of Faculty of Engineering

The Faculty of Engineering was founded in 1991 and was built upon existing strength with added talents from all over the world. The Faculty has been able to attract some of the best minds to the Faculty. Many received their training in leading universities in North America, Great Britain and Australia. Most of them have extensive experience in industry and many are leaders in their fields. This team of top-notch talents is gathered together to nurture local talents through educational programmes, and break new frontier in research through innovative and exciting research endeavours.

The positioning of The Shun Hing Institute of Advanced Engineering in William M.W. Mong Engineering Building is deliberate as a key nucleating point to integrate research endeavours in the Engineering Faculty and its neighboring Faculties. Many members of the Zcience and Medical Faculties are comrades of this Faculty in many interesting research collaborations. It is the ambitious goal of the Faculty of Engineering that the Institute should become a light house for the local technology landscape and will show the directions to push forward the migration towards high value-added technology and information economy.

The mission of the Institute is to spearhead, conduct, promote and co-ordinate research in advanced engineering topics. There is no end to the list of areas to be explored and the plan is to give priority to research topics that are both exciting and innovative. The Institute is also aspired to transfer its research results to industry for practical application and to put across to the community at large the role of engineering as a driving force for human development through educational activities.

Building on Strength and The Way Ahead

The research projects to be pursued will be built upon areas in which the Faculty has already achieved outstanding performance. These are areas that have great potential for further technological advancement and in line with industrial development in Hong Kong. The Institute provides a vibrant R&D environment to spur new discoveries and speed up their translation into applications. Initially, two research centres are established, focusing on research in ***Biomedical Engineering*** and ***Multimedia Technologies***.

Technology Transfer

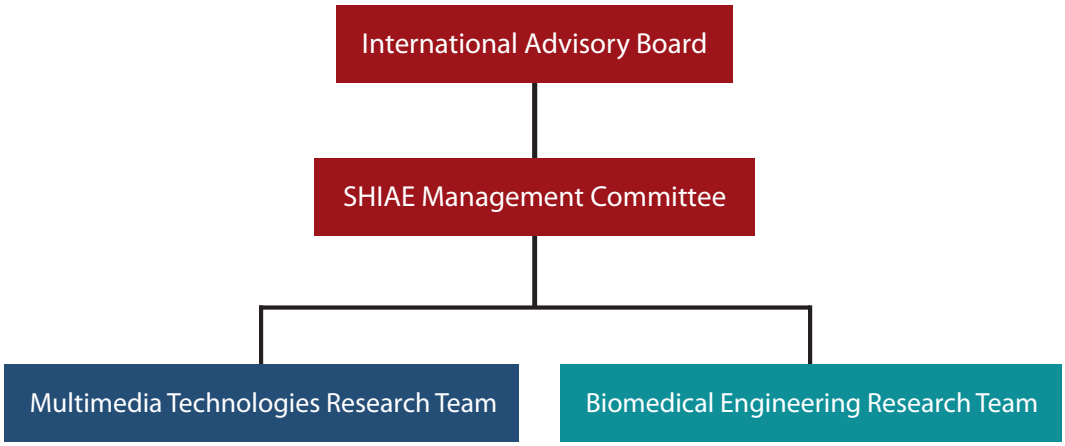
To bring the Institute one step closer to reality, there is a commitment to create a healthy flow of talent and knowledge into industry. Synergy with industry is going to be the ultimate goal to approach research and development in Hong Kong. External experts will be settling in the Institute to lead research projects that could benefit the industrial sector.

The technology transfer arm of the Faculty of Engineering will be playing an important role to go between Institute and industry. When fully operational, the Institute will house an array of top-notch research and development activities encompassing contract research, spin-off companies, and consultancies.

Contribution to Society

The Institute will contribute to the progress of Hong Kong through a wide range of educational activities like training courses, seminars, symposiums to disseminate the latest technologies to promote appreciation of engineering in society and arouse interest of the younger generation towards engineering.

Organization



Composition of International Advisory Board

Chairman :

Mr. David T.Y. Mong 蒙德揚先生

Managing Director
Shun Hing Electronic Trading Co., Ltd.
Hong Kong

Members :

Professor Victor Zue

Professor of Electrical Engineering and Computer Science
Co-Director of the Institute's Computer Science and
Artificial Intelligence Laboratory (CSAIL) MIT
USA

Professor Benjamin W. Wah

Franklin W. Woeltge Endowed Professor
Department of Electrical and Computer Engineering
Research Professor, Coordinated Science Laboratory
University of Illinois, Urbana-Champaign
USA

Dr. Harry Shum 沈向洋博士

Corporate Vice President
Microsoft Corporate
USA

Professor Yongmin Kim

Professor & Chair Bioengineering
Professor of Electrical Engineering
adjunct with Radiology and Computer Science & Engineering
Department of Bioengineering
University of Washington
USA

Professor Nicholas Ayache

Research Director
EPIDAURE Research Group
The French Research Institute of Computer Science and Automatic Control (INRIA)
France

Professor Jack C.Y. Cheng 鄭振耀教授

Pro-Vice-Chancellor
Professor of Orthopaedics and Traumatology
Faculty of Medicine
The Chinese University of Hong Kong
Hong Kong

Professor Peter T.S. Yum 任德盛教授

Dean of Faculty of Engineering
Professor of Information Engineering
The Chinese University of Hong Kong
Hong Kong

Professor Wing W.S. Wong 黃永成教授

Dean of Graduate School
Professor of Information Engineering
The Chinese University of Hong Kong
Hong Kong

Professor Pak Chung Ching 程伯中教授

Pro-Vice-Chancellor
Director, Shun Hing Institute of Advanced Engineering
Professor of Electronic Engineering
The Chinese University of Hong Kong
Hong Kong

Composition of Management Committee

Director :

Professor Pak Chung Ching
Pro-Vice-Chancellor
Professor of Electronic Engineering

Secretary :

Professor John C.S. Lui
Professor of Computer Science and Engineering

Members :

Professor Peter T.S. Yum
Dean of Faculty of Engineering
Professor of Information Engineering

Mr. Jimmy Wang
Deputy Managing Director of Shun Hing Electronic Trading Co., Ltd.
Hong Kong

Professor Dennis Y.M. Lo
Associate Dean (Research) of Faculty of Medicine
Professor of Chemical Pathology

Professor Ronald C.K. Chung
Professor of Mechanical and Automation Engineering

Professor Y.T. Zhang
Professor of Electronic Engineering

Professor Helen M.L. Meng
Professor of Systems Engineering & Engineering Management

Professor Wing W.S. Wong
Dean of Graduate School
Professor of Information Engineering

Core Members of Research Teams

Biomedical Engineering Track:

Faculty of Engineering

Professor Lai-Wan Chan
Professor Pheng-Ann Heng
Professor Aaron H.P. Ho
Professor Philip H.W. Leong
Professor Wei-Hsin Liao
Professor Wen-Jung Li
Professor Emma MacPherson
Professor Max Q.-H. Meng
Professor William S.-Y. Wang
Professor Ke-Li Wu
Professor Yuan-Ting Zhang

Faculty of Medicine

Professor Jack C.Y. Cheng
Professor K.S. Leung
Professor Mary Miu Yee Waye

Multimedia Technologies Track:

Faculty of Engineering

Professor P.C. Ching
Professor Dah-Ming Chiu
Professor Ronald C.K. Chung
Professor Kin-Chuen Hui
Professor Leo Jiaya Jia
Professor Tan Lee
Professor Jack Y.B. Lee
Professor Wen-Jung Li
Professor John C.S. Lui
Professor Michael Rung-Tsong Lyu
Professor Helen M.L. Meng
Professor King-Ngi Ngan
Professor Raymond Yeung

Faculty of Engineering website: http://www.erg.cuhk.edu.hk
Faculty of Medicine website: http://www.med.cuhk.edu.hk

Shun Hing Visiting Scholars / Fellows

The Institute has launched a Shun Hing Distinguished Scholar Program ever since its inception with an aim to attract distinguished scholars to pursue research collaboration with our faculty and to strengthen our research profile. The following scholars have worked either on a short term or on a longer term engagement with the Institute.

Shun Hing Visiting Scholars:

Professor Ning Xi *(2005 & 2006)*
Dept. of Electrical and Computer Engineering at Michigan State University, USA

Professor Tzung K. Hsiai *(2005 & 2006)*
Dept. of Biomedical Engineering, University of Southern California, USA

Professor Steve C.H. Tung *(2005 & 2006)*
Dept. of Mechanical Engineering, University of Arkansas, USA

Professor Bradley J. Nelson *(2005)*
Swiss Federal Institute of Technology, Zurich

Professor William C. Tang *(2006)*
Dept. of Biomedical Engineering, University of California, Irvine, USA

Professor Ke Xu *(2006)*
Dept. of Computer Science, Tsinghua University, China

Professor Towsley Don *(2006)*
Dept. of Computer Science, University of Massachusetts, USA

Professor Yong Chiang Tay *(2005)*
Dept. of Mathematics and Dept. of Computer Science, The National University of Singapore

Professor Zhuoqing Morley Mao *(2006)*
Dept. of Electrical Engineering and Computer Science, University of Michigan, USA

Professor Hui Zhang *(2006)*
School of Computer Science, Carnegie Mellon University, USA

Dr. Peter Key *(2006)*
Microsoft Research, European Research Centre, Cambridge, USA

Professor Miroslaw Malek *(2005)*
Institut für Informatik, Humboldt-Universitaet zu Berlin, Germany

Professor Jason Yi-Bing Lin *(2005)*
Dept. of Computer Science, National Chiao Tung University, Taiwan

Professor Lui Sha *(2007)*
Dept. of Computer Science, University of Illinois at Urbana-Champaign, USA

Professor Ruilian Zhao *(2007)*
Computer Science Dept., Beijing University of Chemical Technology, China

Professor Shivkumar Kalyanaraman *(2006)*
Dept. of Electrical, Computer and Systems Engineering, Rensselaer Polytechnic Institute, USA

Shun Hing Fellows:

Dr. Hongliang Li *(2006 & 2007)*
Dept. of Electronic Engineering, The Chinese University of Hong Kong

Dr. Wentao Gu *(2006 & 2007)*
University of Tokyo, Japan

Dr. Defeng Wang *(2006 & 2007)*
The Hong Kong Polytechnic University, Hong Kong

Dr. Lisheng Xu *(2006 & 2007)*
Harbin Institute of Technology (HIT), China

Biomedical Engineering

Biomedical engineering (BME) including medical devices, medical imaging and virtual surgery are having an enormous impact on healthcare. Almost all aspects of medicine are being enhanced by biomedical engineering and healthcare technologies. Biomedical engineers are increasingly being involved with the development of novel, reliable, noninvasive technologies for the diagnosis, monitoring, and treatment of diseases. For examples, medical electronic devices and biosensors improve home/mobile healthcare; microarray technologies offer immense potential for studying the whole cellular transactions; imaging tools enhance diagnosis; robotic devices facilitate surgery and rehabilitation; visualization and virtual reality provide excellent means for surgery training and planning.

Since 1999, the Biomedicine research teams in the Faculty of Engineering have acquired more than HK$50 millions form Hong Kong Research Grants Council and Innovation and Technology Fund in related research areas. Highly functional prototypes including wearable medical devices, virtual bronchoscopy, virtual acupuncture, virtual arthroscopy, as well as a web-based scoliosis clinical multimedia database have been developed in clinical use at Prince of Wales Hospital. There is also a very strong and productive collaboration among the engineering and medicine faculties at CUHK in the area of biomedical engineering.

Several significant BME projects have been carried out successfully at the Chinese University of Hong Kong, including Wearable Intelligent Sensors and Systems for e-Health (WISSH) and the Chinese Visible Human Project, in which several ultra-high resolution digital human datasets were collected within the last two years. The team of the Chinese Visible Human Project succeeded in handling massive image processing, 3D reconstruction as well as developed the world fastest photo-realistic volume visualization tools to support stereoscopic rendering of these gigantic digital human datasets.

Specific projects to be pursued include:

1. Biosensors and Medical Devices (BMD): To meet the huge market needs, the focus of this topic will be placed on the up-stream research on new principle of wearable devices and biosensors for the monitoring, diagnosis, and treatment of chronic diseases and for aging population.

2. Tele-Medicine and Mobile Home Healthcare (TM^2H^2) Systems: Under this topic, we propose to develop medical robots and easy-to-use man-machine or brain-machine interfaces suitable to facilitate and promote the use of the wearable biosensor-based medical devices in a larger scale beyond hospitals, and to integrate all the technologies above for the development of the TM^2H^2 Systems that can be widely used for distributed diagnosis and home healthcare.

3. Medical imaging, virtual anatomy and virtual surgery: We aim to develop advanced computational tools and engineering techniques to analyze, visualize, manipulate and distribute functional imaging of the human brain in health and disease, in order to achieve better understanding of neuro-diseases and to enhanced computer-assisted diagnosis. We also aim to construct next generation virtual environments that facilitate interactive learning of functional neuroanatomy and dexterous training of neuro-endoscopic surgeries.

Research Reports in **Biomedical Engineering**

Completed Projects

- **Cochlear Implants**
- **Virtual Anatomy and Dexterous Simulators for Minimal Access Cardiothoracic and Neuro-endoscopic Surgeries**
- **Systematic Synthesis of Nano-informatics Chips by Nano-Robotics Manipulation**

On-going Projects

- **Study on RF Radiation Effect and Signal Efficiency of Wireless Medical Devices**
- **Photonic Biosensor Micro-arrays for Molecular Diagnostics Applications**
- **Bio-electromagnetic Modeling and Experiment Setup for Medical Electronics RF Safety Assessment**
- **Medical Applications of Terahertz Imaging**
- **Hybrid Assistive Knee Braces with Smart Actuators**
- **Research on Language and Brain Waves**

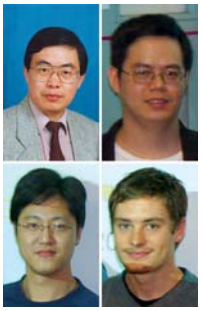
Cochlear Implants

Principal Investigator:

Professor Yuan-Ting Zhang [1]

Research Team Members:
Dr. Fei Chen and Dr. Tian Guan, Postdoctoral Fellows [1]
Mr. Florent Cosandier, Research Assistant [1]

[1] *Dept. of Electronic Engineering, CUHK*



Project Start Date: 1st March 2005
Completion Date: 31st March 2007

Abstract

Nowadays, cochlear implant (CI) has been widely accepted as the most useful solution for the treatment of severe-to-profound hearing loss. CI employs direct electrical stimulation of the auditory nerves to restore some degree of hearing. Early work has shown that although present CI devices produce a better speech perception to English users, many people speaking tonal languages show poor speech perception results. Such performance variation is believed to be mainly due to the following reasons: 1) CI users gain little tonal information from present electrical stimulation strategies, although it has been well known that the tonal language uses different tones to indicate diverse lexical meanings; and 2) the neural firing pattern from electrical-stimulation-based electrical hearing varies from that of normal acoustic hearing. In order to enhance the speech perception of CI users who speak Chinese languages, including Mandarin and Cantonese, we developed a novel auditory-model-based electrical stimulation strategy, which incorporated the tonal information and the neural discharge pattern mimicking that in normal human ear.

1. Objectives

1) To study bio-model-based time-varying signal processing techniques for reinforcing intonation of speech;
2) To propose an innovative signal processor for CI catered not only for the English language but also two Chinese dialects, Mandarin and Cantonese; and
3) To develop an auditory-model-based electrical stimulation strategy incorporating tonal information for cochlear implants.

2. A Stochastic Model of Electrically Stimulated Auditory Nerve

Since the electrical pulses straightforwardly stimulate the auditory nerve (AN) in CI, modeling the AN's response to electrical stimulation would improve our understanding on how auditory percepts are produced by CI, and facilitate our development of new electrical stimulation strategies. A simple but efficient integrate-and-fire based AN model was proposed, as shown in Fig.1. The model employed a Gaussian noise and a refractory function to describe the stochastic membrane potential fluctuation and the refractory effect during neural firing, respectively [2]. The proposed model was characterized by its computation ease and efficiency. Based on the model, we conducted the following two studies to investigate the AN response to electrical stimulation.

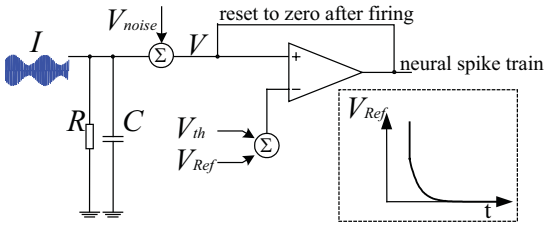


Fig.1. Integrate-and-fire based stochastic model of electrically stimulated auditory nerve. V_{Ref} denotes the refractory function for neural threshold.

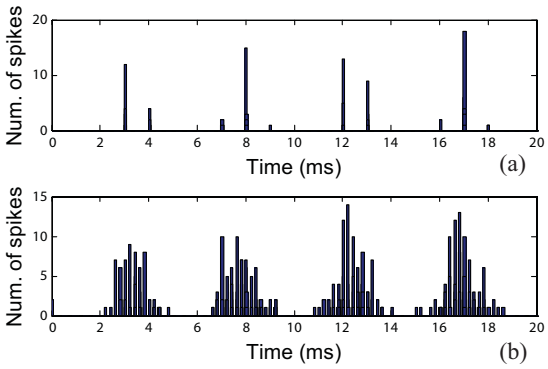


Fig.2. PST histogram of AN model response to electrical stimulation with pulse rate of (a) 1000 pps and (b) 5000 pps. The modulating waveform is a 220 Hz sinusoid.

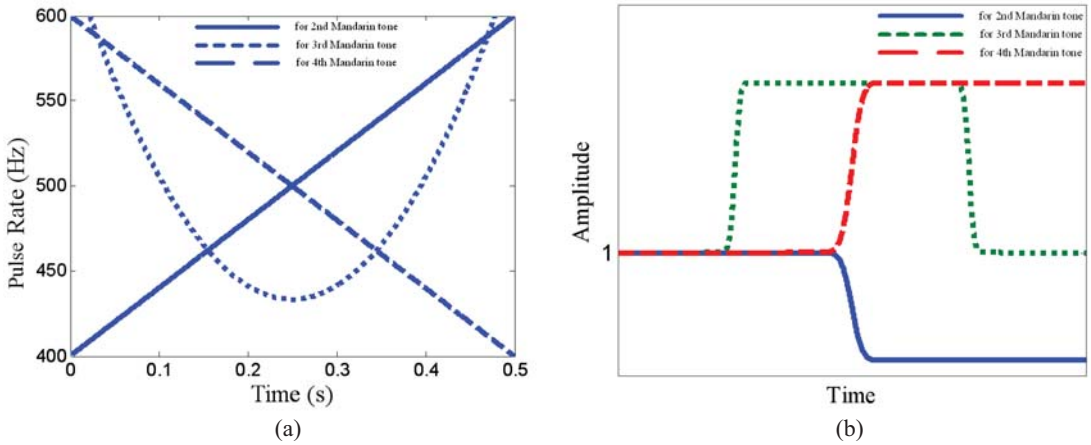


Fig.3. (a) The modulated pulse-rates, and (b) the associated amplitude compensation schemes to convey the 2nd, 3rd, and 4th Mandarin tonal patterns, respectively.

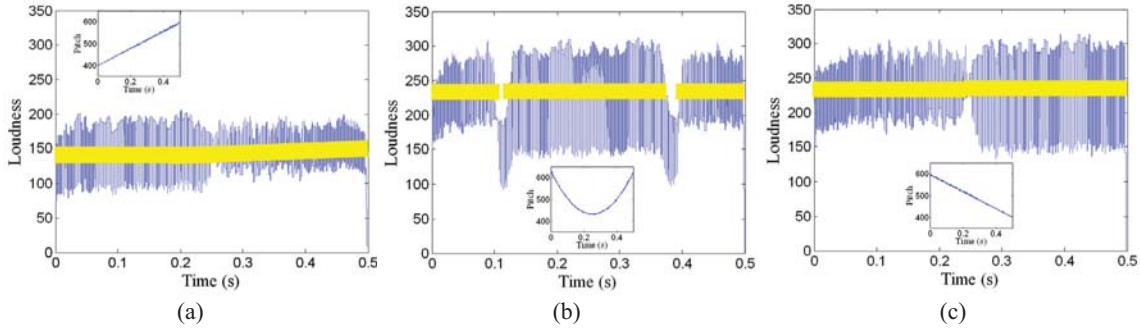


Fig.4. The estimated loudness patterns after amplitude compensation for electrical pulsatile stimulation using pulse-rate modulation to convey the (a) 2nd, (b) 3rd, and (c) 4th Mandarin tonal patterns, respectively. The inserted plot in each figure indicates the corresponding estimated pitch pattern.

2.1. Response to High-rate Electrical Pulse Train to Encode the Temporal Fine Structure

There have been growing evidences that electrical stimulation strategies using low or medium pulse rate would cause the synchronization of AN's firing activity. The synchrony subsequently leads to a loss of temporal resolution in that multiple fibers code the timing of the peak of a stimulus rather than its entire waveform. Fig.2 (a) illustrated this limitation with the post-stimulus time (PST) histogram of a 220 Hz sinusoid by using an electrical stimulation with a rate of 1000 pulse per second (pps). Nevertheless, when a high electrical stimulus rate 5000 pps was utilized, a rather different PST histogram result was demonstrated in Fig.2 (b), which dramatically improved the temporal resolution of the simulated fibers and allowed them to encode most attributes of the temporal waveform.

2.2. Auditory-nerve-model-based Loudness Normalization Strategy for CI Electrical Stimulation

Recent psychological experiments have found that the increment of pulse-rate would increase the loudness sensation perceived by CI users, and cause some detrimental effects on speech perception. Besides, considering the limited dynamic hearing range of CI users, it is necessary to diminish such synthetic loudness variation or normalize the loudness perception.

Based on the proposed AN model, we introduced an amplitude compensation scheme to normalize the loudness perception when the pulse-rate was modulated to convey Mandarin tonal information [3]. When the pulse-rate was modulated according to different Mandarin tonal patterns as shown in Fig.3 (a), a specific time-varying function was used to adjust the amplitude of the pulsatile train accompanying the pulse-rate modulation, as shown in Fig.3 (b). A model-based study was conducted to investigate the performance of loudness normalization. Results of the model-based simulation in Fig.4 showed that using the proposed amplitude compensation scheme, the estimated loudness was normalized while the Mandarin tonal information could still be efficiently transmitted. Therefore, it is believed that, when the proposed electrical pulsatile stimulation incorporating both pulse-rate modulation and amplitude compensation is integrated with present CI devices, it would more efficiently enhance the speech perception, both pitch and loudness perceptions, for cochlear implantees speaking tonal languages.

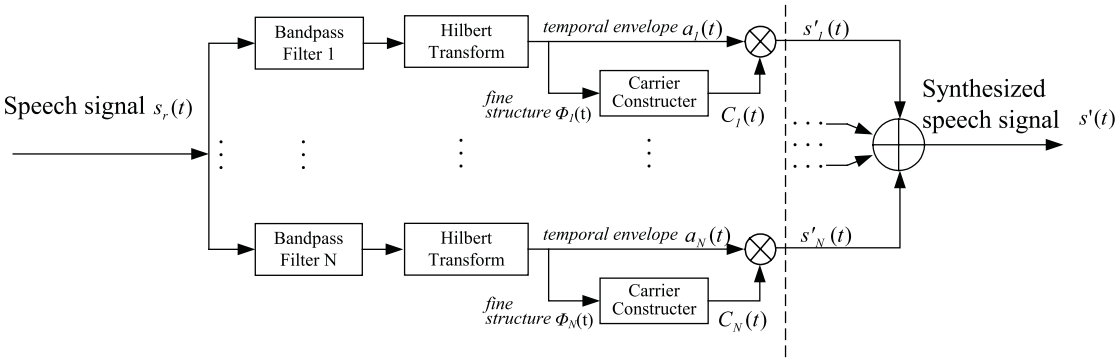


Fig.5. The block diagram of the proposed speech decomposition/synthesis model incorporating the FS cue.

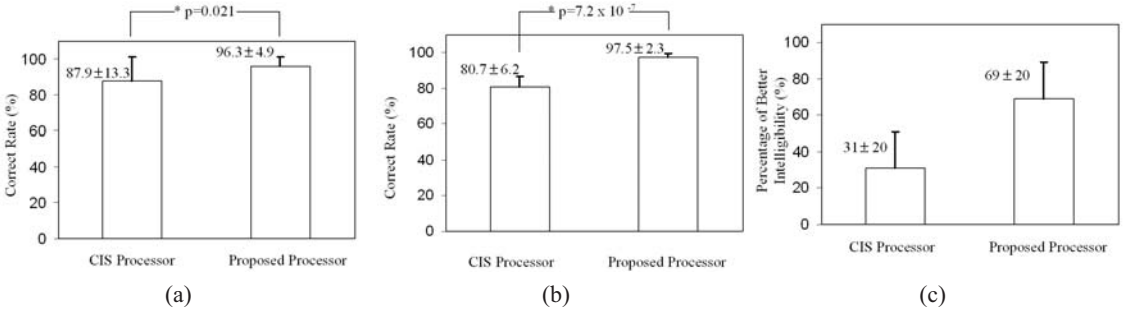


Fig.6. The correct rates of (a) Mandarin tone identification and (b) speaker recognition, and (c) the percentages of better intelligibility of the synthesized sounds by CIS processor and the proposed processor. Asterisk (*) identifies significant difference between two conditions.

3. A Fine-structure-based Speech Processor for Cochlear Impants

According to Hilbert transform (HT), a signal can be decomposed into a slowly-varying envelope modulating a high-frequency carrier, or fine structure (FS), of the original signal. Temporal envelope and fine structure have been recognized as two important acoustic cues for speech intelligibility. Most of the current CI speech processors, such as the well-known continuous-interleaved-sampling (CIS) processor, basically emphasize the envelope cue, based on the fact that using the slowly-varying temporal envelope extracted from only 3 to 4 frequency bands might produce a nearly perfect speech recognition in the quite environment. However, fine structure has been recently found to contain more acoustic cue for pitch recognition than temporal envelope. Currently, studies on how to efficiently make use of the FS cue are actively ongoing, and may significantly influence the design of the electrical stimulation strategy of CI in the near future.

We proposed a speech decomposition/synthesis model, which incorporated the FS cue of the speech signal in order to enhance the pitch perception of CI users [4-5]. The model structure is illustrated in Fig.5. After bandpass filtering the speech signal into multiple frequency bands, temporal envelope and fine structure are extracted by using HT. A carrier signal is constructed for each band by using a train of high-rate sinusoidal pulses located at the peak positions of the fine structure. The carrier signal is then amplitude-modulated by the temporal envelope in the band so as to produce the band-specific decomposition output. Speech synthesis is implemented by summing the decomposition outputs from all the bands.

3.1. Acoustic Simulation Experiments

Acoustic simulation experiments were conducted to investigate the contribution of the model-based speech processor for Mandarin tone identification, speaker recognition, and subjective assessment of speech intelligibility.

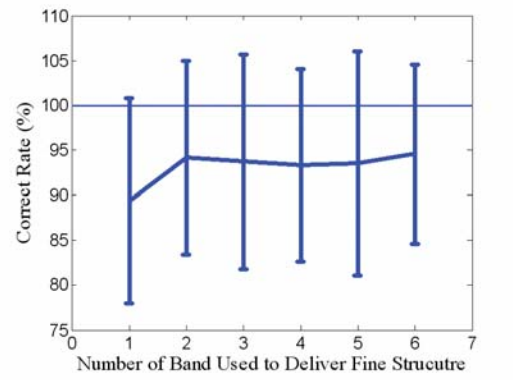


Fig.7. The correct rates of Mandarin tone identification by using FS cues from different amount of LF bands.

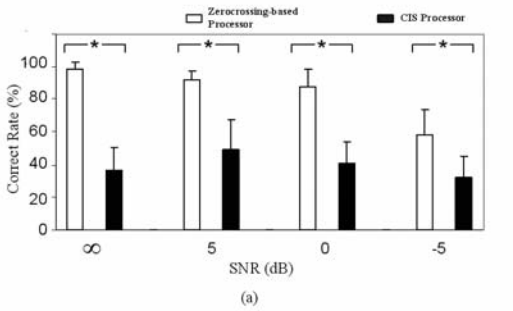
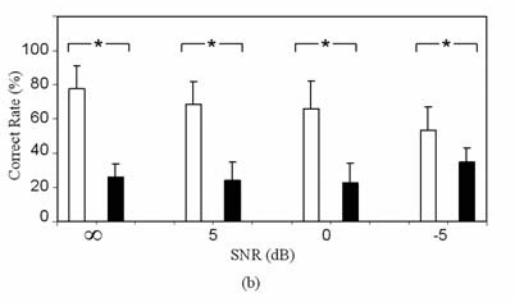
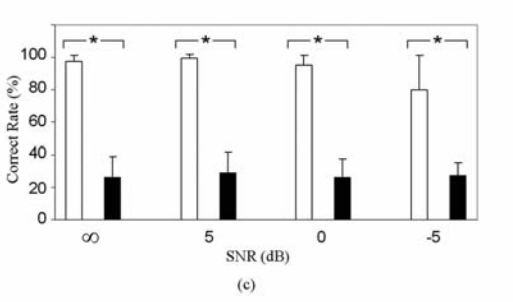


Fig.8. The correct rates of Mandarin tone identification as a function of SNR when the speech signals are decomposed into (a) 4, (b) 6 and (c) 8 bands, respectively.





3.1.1. Mandarin Tone Identification

We collected two sets of speech signals. The first set had 100 Mandarin speech signals, which included 25 consonant-vowel combinations with each having four Mandarin tonal patterns. 12 Mandarin-speaking subjects participated in the experiment of Mandarin tone identification. Each subject listened to 40 sounds synthesized by CIS processor and the proposed processor in their 6-band versions, respectively. As shown in Fig.6 (a), the correct rates for Mandarin tone identification were 87.9 13.3% and 96.3 4.9% by using CIS processor and the proposed processor, respectively.

3.1.2. Speaker Recognition

The second speech set included Mandarin phrases recorded from 8 native Mandarin-speakers (4 males and 4 females). Ten additional listeners (5 males and 5 females) participated in the experiment of speaker recognition. Fig.6 (b) showed the correct rates of speaker recognition by the two processors. The correct rates for CIS processor and the proposed processor were 80.7 6.2% and 97.5 2.3%, respectively.

The paired *t*-test results in Fig.6 (a) and (b) indicated that there were significant differences between the two processors for their performances in Mandarin tone identification and speaker recognition.

3.1.3. Subjective Assessment of Speech Intelligibility

Fig.6 (c) showed the percentages of better intelligibility of the synthesized voices by the two processors. 31% of the voices synthesized by the CIS processor sounded more intelligible, while 69% by the proposed processor were recognized with better intelligibility.

Furthermore, we conducted two studies for the efficient implementation of the proposed FS-based CI speech processing strategy.

3.2. Frequency-specific FS Importance for Mandarin Tone Perception

The first study attempted to explore the significant frequency bands for conveying the FS cue [6]. The speech signal was split into six frequency bands by band-pass filtering. After obtaining the fine structures of the speech signals in each band by using HT, only the fine structure was transmitted to 12 Mandarin-speaking subjects. Six algorithms were designed to utilize the frequency-specific FS cues. Algorithm one only contained the FS cue of the lowest frequency band, and each of the remained algorithms cumulatively incorporated one more high-frequency (HF) band.

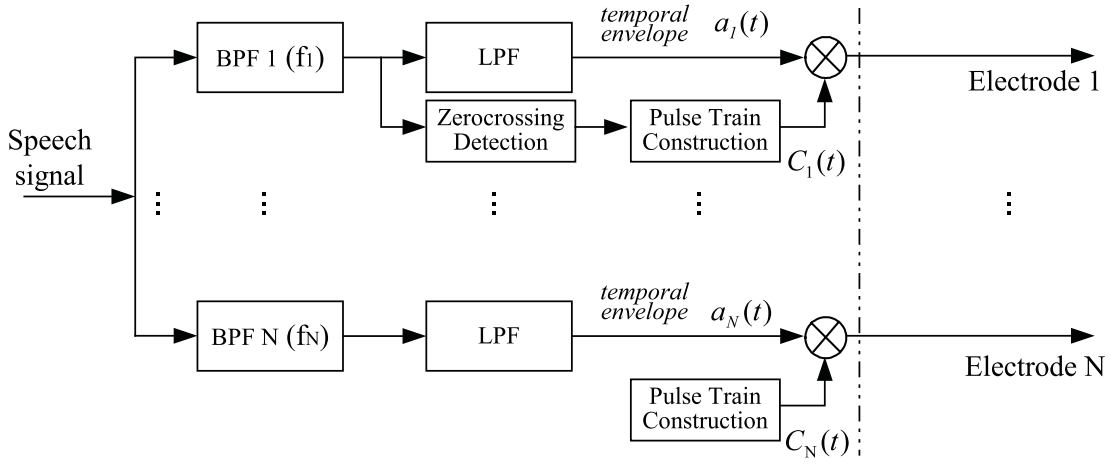


Fig.9. The block diagram of the proposed electrical stimulation strategy incorporating tonal information.

The result in Fig.7 showed that: 1) when only the FS cue from the lowest frequency band was conveyed to the listener, the listener achieved a tone recognition rate around 90%; 2) the tone recognition rate increased to 95% when FS cues from two low-frequency (LF) bands were used. Further recruitment of more fine structures from HF bands did not significantly improve the tone recognition rate. Therefore, the experimental results indicated that the FS cue from the LF band of speech signal would be able to efficiently convey Mandarin tonal information.

3.3. Zerocrossing-based FS Representation to Convey Mandarin Tonal Information

The second study aimed to alleviate the computation complexity of the proposed model by using a simple zerocrossing-based method to represent the FS cue [7]. The high-rate sinusoidal pulses were located at the zerocrossing positions of the speech signal after band-pass filtering. An acoustic simulation experiment was conducted to evaluate its effectiveness on Mandarin tone identification. Eight Mandarin-speaking subjects listened to the monosyllabic sounds synthesized by CIS processor and the proposed zerocrossing-based processor.

Fig.8 showed the performances of Mandarin tone identification by CIS processor and the zerocrossing-based processor. When there was no noise contamination, the correct rates of CIS processor were 36.7%, 25.8% and 25.8%, while those of the zerocrossing-based processor were 98.3%, 77.5% and 97.5% when the speech signals were decomposed into 4, 6 and 8 bands, respectively, indicating that the zerocrossing-based speech processor performed better than CIS processor in improving Mandarin tone identification. Furthermore, when the speech-spectrum shaped noise was added to the stimulus at three SNR levels, i.e. 5dB, 0dB and -5dB, the zerocrossing-based processor consistently produced higher correct rates than CIS processor in the three noisy conditions.

4. Electrical Stimulation Strategy Incorporating Tonal Information for Cochlear Implants

Based on the FS-based speech synthesis model, and motivated by the frequency-specific importance of FS cue and the effectiveness of zerocrossing based FS representation, a new CI electrical stimulation strategy was developed, as shown in Fig.9 [10]. The zerocrossing times in the LF bands were used to represent the FS cue in those bands, and to trigger a pulse at the zerocrossing time. The HF bands utilized the traditional interleaved electrical pulse generation scheme, as CIS strategy. To avoid the possibility of two simultaneous pulses between LF and HF bands, the lower band was given a higher priority for electrical stimulation.

Acoustic simulation experiment was conducted to investigate the performance of the proposed stimulation strategy. Eight Mandarin-speaking subjects listened to the monosyllabic sounds synthesized by CIS and the proposed zerocrossing-based electrical stimulation strategies in their 6-band versions. Fig.10 showed the correct rates for CIS and the proposed strategies when there was no additional noise, and the speech-spectrum shaped noise was added to the stimulus at three SNR levels, i.e. 5dB, 0dB and -5dB, respectively. Consistent higher correct rates resulted from the proposed electrical stimulation strategy than the traditional CIS strategy, which indicated that the proposed stimulation strategy was able to enhance the deliver of FS cue from speech signal.

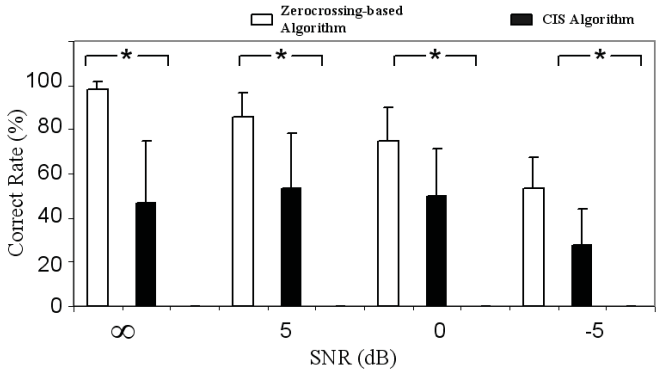


Fig.10. The correct rates of Mandarin tone identification as a function of SNR when the speech signals are decomposed into 6 bands.

5. Summary

An integrate-and-fire based model was developed for the auditory nerve to investigate its response to electrical stimulation. The model supported that an enhanced temporal neural representation of the stimulus waveform could be produced by using the high-rate electrical stimulation. Based on the AN model, a simple and efficient amplitude compensation scheme was introduced to normalize the loudness perception when the pulse-rate of electrical pulsatile stimulation was modulated to convey the Mandarin tonal information. Result of model-based simulation showed that using the proposed amplitude compensation scheme, the estimated loudness was normalized while the Mandarin tonal information could still be efficiently transmitted.

A new CI speech processing strategy based on a speech synthesis model incorporating the temporal fine structure cue was developed. The model used the zerocrossing times to represent the FS cue straight from the temporal domain, and represented it in a way analogous to the pulsatile electrical stimulation strategy in present CI devices. The experimental results from the acoustic simulation showed that the model-based speech processor produced significant improvement on several subjective assessments associated with tone perception.

Based on the experimental indication on the frequency-specific importance of FS information, a new CI electrical stimulation strategy was developed. The strategy was able to deliver the tonal information; meanwhile, it did not introduce inter-channel crosstalk during electrical stimulation.

6. Publications and Invited Talk

Invited Talk

[1] "Speech Processing Strategy Incorporating Tonal Information for Chinese Cochlear Implantees", Expert Meeting 2007: A Cochlear Implant System for China, held in Prince of Wales Hosptial, The Chinese University of Hong Kong, May 4-5, 2007.

Publications

[2] "An integrated-and-fire based auditory nerve model and its response to high-rate pulse train," *Neurocomputing*, vol. 70 (4-6), pp.1051-1055, Jan. 2007.

[3] "Loudness normalization for cochlear implant using pulse-rate modulation to convey Mandarin tonal information: A model-based study," in *Proc. 28th Ann. Int. Conf. IEEE-EMBS*, New York, 2006, pp.1236-1239.

[4] "A novel temporal fine structure based speech synthesis model for cochlear implant," submitted to international journal.

[5] "A new acoustic model incorporating temporal fine structure cue for cochlear implant," in *Proc. Int. Conf. Info. Tech. Appl. in Biomed.*, Greece, 2006.

[6] "Frequency-specific fine structure cues from Mandarin for designing electrical stimulation strategy of cochlear implant," in *Proc. World Cong. on Med. Phy. and Biomed. Eng.*, Seoul, 2006, pp. 5068.

[7] "Zerocrossing-based representation of the temporal fine structure: Towards improving the tone identification for cochlear implantees," in *Proc. BME 2006 Biomed. Eng. Conf.*, Hong Kong, 2006, pp.89-92.

Virtual Anatomy and Dexterous Simulators for Minimal Access Caradiothoracic and Neuro-endoscopic Surgeries

Principal Investigator:

Professor Pheng-Ann Heng [1]

Co-investigators:

Professor T.T. Wong [1], Professor Jack C.Y. Cheng [2], Professor K.S. Leung [2], Professor C.W. Chu [3], Professor W.M. Lam [3], Professor W.S. Poon [4], Professor David T.W. Yew [5]

Research Team Members:

Mr. Chui Yim Pan, Research Associate [1]
Dr. Xie Jun, Postdoctoral Fellow [1]
Dr. Wang Defang, Postdoctoral Fellow [1]
Ms. Wang Xinghe, Research Assistant [1]
Ms. Qu Yingge, PhD student [1]
Ms. Shi Lin, PhD student [1]
Mr. Xie Yongming, MPh student [1]

[1] Dept. of Computer Science and Engineering, CUHK
[2] Dept. of Orthopaedics and Traumatology, CUHK
[3] Dept. of Diagnostic Radiology and Organ Imaging, CUHK
[4] Dept. of Surgery, CUHK
[5] Dept. of Anatomy, CUHK

Project Start Date: 1st March 2005
Completion Date: 31st October 2007

Abstract

Building upon our recent advanced research development in virtual reality, visualization, imaging and biosensors research, together with our medical expertise and our strong partnership in the Chinese Visible Human Project, this research team has the unique combined strength and adequate imaging resource to develop next generation highly interactive and realistic virtual environments for medical education and surgical training. The latest CVH male dataset is about 1143 GB in size. There are in total 18,200 cross-section digital images (4064x2704 pixel resolution, 48 bits color) at 0.1mm intervals; extremely fine anatomical details can be observed. We would like to explore new possibilities for using such ultra-high resolution data in multiple directions for medical research, education, surgical training and simulation.

In the project, our long-term key **objectives** include:

1) Through original research, innovative application and intelligent integration of new advances in computing and engineering to develop state-of-the-art virtual anatomy and surgical simulators.
2) To advance the frontier of medical visualization, biomedical computing and digital human research.
3) To promote VR-based medical education and surgical training in China and the region.

Through this proposed research, we would like to develop highly interactive virtual environments that support in-depth learning of virtual and functional anatomy, focusing on cardiac and neuro anatomy.

The significance and long-term impacts of the proposed project are as follows:

- The proposed project will contribute significantly to the basic and applied research in the fields of digital human visualization, medical virtual environment and virtual human anatomy. It would also generate practical integrated VR-based systems for learning cardiac anatomy and neuroanatomy, as well as surgical training in minimal access neuro-endoscopic surgeries.

- Our deliverables can enhance and facilitate computer based teaching and learning in medicine. Computing technology is ubiquitous in today healthcare environments; it is critical to expose our clinicians and surgeons to these kinds of technologies and to take full advantages of such engineering advancements in their training and research.

1. Research Methodology

We shall carry out original research in the following tasks in order to achieve our goal:

1.1. PC-based real-time photo-realistic rendering based on GPU

Only with the recent development in graphics accelerator, it becomes possible to achieve real-time multi-gigabyte volume visualization on average PC. The visualization challenge in this project is to real-time render extremely large digital human dataset. We have achieved break-through development in using hybrid 2D-3D texture rendering approach to realize PC-based interactive visualization of Visible Human Male dataset (15 GB). It would be a great challenge to realize interactive visualization of the recent Chinese Visible Human datasets (90GB, 141GB, 1143GB). Furthermore, with the additional requirement of motional and functional simulations, even more speedy approach would be needed. We would like to exploit further the latest programmable graphics hardware features combine with optimal usage of the built-in texture memory and intelligent clustering technology.

1.2. Construction of functional virtual organ models

With the latest development in visible human research, it is now a perfect time to create a unified whole-body system of spatial coordinates to which all medical work can be referred; we have sufficient expertise and the imaging resource to construct such a virtual human, but it is a longer term effort and is beyond the scope of this project. Nonetheless, we shall construct organ-by-organ model of the virtual human, focusing on the neuroanatomy for a start. These organ models will be interactively modifiable, using a 3D interface, for learning anatomy and planning therapy.

1.3. Illustrative visualization for anatomy learning

Special visualization features will be supported in our virtual environments. For example multi-layer visualization, which is peeling of skin to the fat layer to reveal the muscle layer, following by further peeling to reveal the inner organ layer and then continue to the skeleton layer. Also we would like to include motional and functional features, for example skeleton movements, blood

circulation, etc. One superior area of this new Chinese visible human data set as compare to the original US visible human data set is the clear differentiation of the blood vessel structures in the color cross-sectional images. Even that, it is still a challenge to extract the complete vascular system form such a huge dataset.

1.4. Multi-sensory feedbacks in the virtual environment

In order to perform interactive analysis of anatomical structures, we will develop a simple, intuitive, yet versatile user interface for changing the mode of display and performing interactive volumetric anatomy visualization. Additional sensory feedback such as tactile will be used to provide a realistic and user-friendly virtual environment. Much emphasis will be placed on the design of effective human-machine interface and requires substantial amount of interfacial research and technology development. To achieve a most realistic virtual environment possible, we shall address the following topics:

- The effective use of stereo for enhancing depth cueing
- The integration of haptic interface for interactive manipulation of the virtual organs
- Dexterous virtual control within a reach-in virtual space
- High performance flexible visualization tools that facilitate comprehensive medical data analysis and data manipulation
- Interface technology for refined multimodal command, control and appropriate multimedia responses

1.5. System integration

Intelligent integration and flexible customization of various technologies developed from the above research areas are critical in order to provide ideal platforms and effective virtual environments for supporting high-tech medical education and surgical training.

2. Results Achieved

Altogether 11 journal papers, 3 book chapters, and 19 conference papers relevant to this project have been generated. We have also developed a prototype on virtual neuroanatomy as shown in Figure 1. A summary of major findings are given under the following categories:

2.1. Medical image segmentation and analysis

- In [32, 33], we presented a unified approach for segmenting and tracking of the Chinese visible human dataset based on level set approach.
- In [2, 15], we proposed algorithms for fMRI image analysis.
- In [3, 4, 5, 14, 16, 17, 18], we proposed algorithms for morphometric analysis based on MRI skull imaging.
- In [6, 28, 29], we proposed algorithms for cardiac image analysis and motion decomposition.
- In [10, 11, 12], we proposed algorithms for image analysis on color image and ultrasound imaging.

2.2. GPU-accelerated rendering and PPU-accelerated surgical simulation

- In [7, 9], we proposed the use of PPU for accelerating simulation of orthopedics surgery and interventional procedures. The trainer can simulate movement and manipulation of skeletal structures, soft tissue and vessels with convincing deformation and realistic bleeding.
- In [31], we propose the GPU-friendly marching cubes for visualizing translucent iso-surfaces. This powerful algorithm can generate multiple-layer translucent iso-surfaces without performing computational expensive sorting.

2.3. Integrated haptic and visual interface

- In [22, 23], a multi-dissected layers model that can be deployed for the haptic training of needle puncturing is presented.

2.4. VR-based surgical simulation

- In [1], a surgical simulation system for neuro-endoscopy is described.
- In [8, 13], we proposed frameworks for enhancing collaboration in VR-based surgical simulation
- We presented our developed imaging technologies for orthopedics visualization and simulation in [19].
- In [20, 21], we described our developed virtual reality systems for anatomic visualization and orthopedics training.

2.5. Virtual Anatomy

- In [25, 26], we presented our research on anatomical exploration and photorealistic virtual anatomy.
- In [30], we proposed an algorithm for annotating anatomical structure automatically.

2.6. Visible Human-based virtual Medicine

- In [24, 27], we presented our on-going virtual medicine research based on Chinese visible human dataset, such as virtual anatomy and virtual acupuncture.

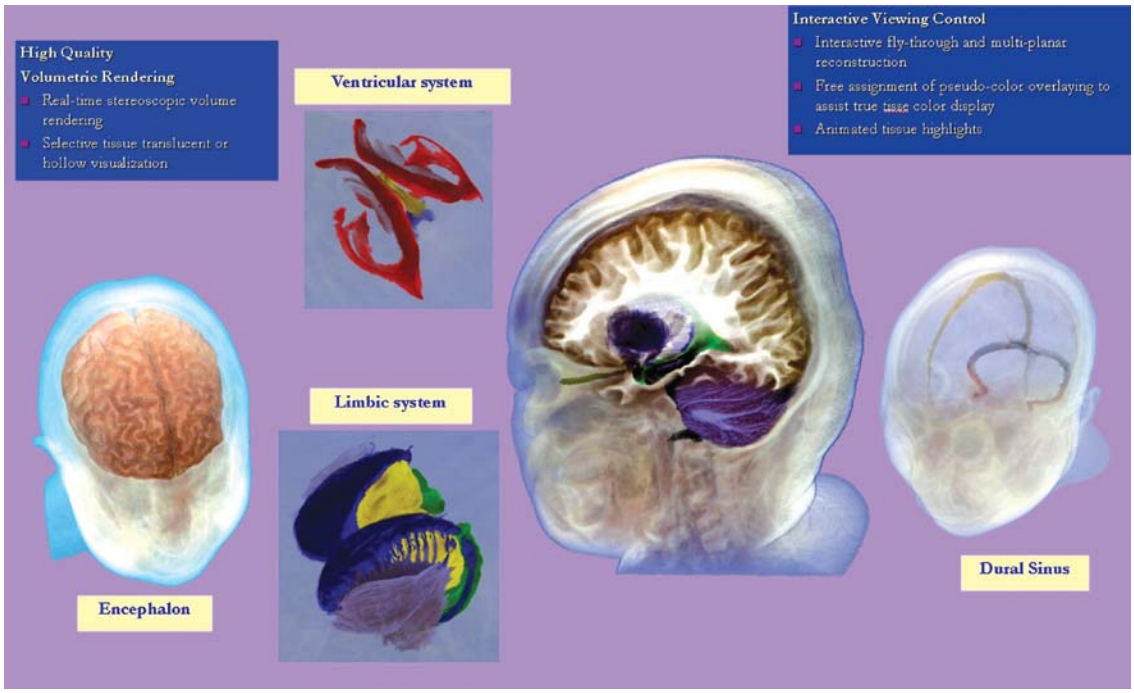


Figure 1. The user interface of our virtual neuroanatomy software.

3. Publications and Awards

3.1. Awards

1. GOLD AWARD WINNER, Hong Kong IT Excellence Awards, Oct 2005, Hong Kong

2. WINNER (R&D category), Asia Pacific Information and Communication Technology Awards, Feb 17 – 19, 2006, Chiangmai, Thailand

3.2. Publications

[1] C. Y. Tang, W. Chin, Y. P. Chui, W. Poon and P. A. Heng, "A Virtual Reality-based Surgical Simulation for Virtual Neuroendoscopy," in Proceedings of *IEEE International Conference on Integration Technology (IEEE ICIT 2007)*, pp. 253 – 258, March 20-24, Shenzhen, China.

[2] D. Wang, L. Shi, D. S. Yeung, C. C. Tsang and P. A. Heng, "Ellipsoidal Support Vector Clustering for Functional MRI Analysis," *Pattern Recognition*, Vol. 40 (2007) pp. 2685 – 2695.

[3] D. Wang, L. Shi and P. A. Heng, "Radial Thickness Calculation and Visualization for Volumetric Layers," in *Proceedings of Open Source and Open Data for MICCAI, a MICCAI 2007 Workshop,* pp. 24 – 31, Nov 2, 2007, Brisbane Australia.

[4] D. Wang, L. Shi and P. A. Heng, "Simultaneous Alignment and Landmark Labelling of Shapes by Minimizing the Description Length," *International Journal of Computer Assisted Radiology and Surgery, (Proceedings of CARS 2007)* Vol. 2- Supplement 1, pp. S118 – S119, June 27-30, 2007, Berlin, Germany.

[5] D. Wang, L. Shi and P. A. Heng, "The Ellipsoidal Harmonic Representation and Its application in Shape Analysis of 3D Anatomical Structures," *International Journal of Computer Assisted Radiology and Surgery, (Proceedings of CARS 2007)* Vol. 2- Supplement 1, pp. S464, June 27-30, 2007, Berlin, Germany.

[6] G. Luo and P. A. Heng, "LV Shape and Motion: B-Spline Based Deformable Model and Sequential Motion Decomposition," *IEEE Transactions on Information Technology in Biomedicine*, Vol. 9, No. 3, pp. 430 – 446, September 2005. (IF = 1.575)

[7] J. Guo, S. Li, Y. P. Chui, Q. Meng, H. Zhang, C. H. Yu and P. A. Heng, "PPU-based Deformable Models for Catheterisation Training," to appear in *Proceedings of Computational Biomechanics for Medicine II, a MICCAI 2007 Workshop*, Oct 29, 2007, Brisbane, Australia.

[8] J Qin, P. A. Heng, K. S. Choi and Simon S. Z. Ho, "An Adaptive Framework using Cluster-based Hybrid Arrchitecture for Enhancing Collaboration in Surgical Simulation," in proceedings of *MMVR 15 (Medicine Meets Virtual Reality 2007)*, pp. 367 – 372, Feb 6 – 9, 2007, Long Beach, California, USA.

[9] J. Qin, W. M. Pang, Y. P. Chui, Y. M. Xie, T. T. Wong, W. S. Poon, K. S, Lueng, and P. A. Heng, "Hardware-accelerated Bleeding Simulation for Virtual Surgery," to appear in *Proceedings of Computational Biomechanics for Medicine II, a MICCAI 2007 Workshop*, Oct 29, 2007, Brisbane, Australia.

[10] J. Xie and P. A. Heng, "Shape Modeling Using Automatic Landmarking," in *Proceedings of Medical Image Computing and Computer-Assisted Intervention MICCAI 2005, LNCS 3750*, pp. 709 – 716, Palm Spring, CA, USA, Oct 26 – 29 2005.

[11] J. Xie and P. A. Heng, "Color Image Diffusion Using Adaptive Bilateral Filter," in *Proceedings of 27th Annual International Conference of the IEEE Engineering In Medicine and Biology Society (EMBS), Shanghai*, September 1-4, 2005.

[12] J. Xie, Y. Jiang, H. T. Tsui and P. A. Heng, "Boundary Enhancement and Speckle Reduction for Ultrasound Images via Salient Structure Extraction," *IEEE Transactions on Biomedical Engineering*, Vol. 53, No. 11, November 2006, pp. 2300-2309.

[13] K. L. Chong, S. W. Tang, J. Qin, Y. P. Chui and P. A. Heng, "ECiSS: A Middleware Based Development Framework for Enhancing Collaboration in Surgical Simulation," in *Proceedings of IEEE International Conference on Integration Technology (IEEE ICIT 2007)*, pp. 15 – 20, March 20-24, Shenzhen, China.

[14] L. Shi, D. Wang, P. A. Heng and T. T. Wong, "Outlier Reduction for Statistical Shape Analysis," *International Journal of Computer Assisted Radiology and Surgery, (Proceedings of CARS 2007)* Vol. 2- Supplement 1, pp. S119 – S120, June 27-30, 2007, Berlin, Germany.

[15] L. Shi, P. A. Heng and T. T. Wong, "Detecting Brain Activation in Functional MRI Data via Kernelized Clustering," in *Proceedings of the 12th International Conference on Biomedical Engineering (ICBME 2005)*, December 7 – 10, 2005, Singapore.

[16] L. Shi, D. Wang, P. A. Heng, T. T. Wong, C. W. Chu, H. Y. Yeung and C. Y. Cheng, "Landmark Correspondence Optimization for Coupled Surfaces," in *Proceedings of MICCAI 2007*, pp. 818 – 825, Oct 29 – Nov 2, Brisbane, Australia.

[17] L. Shi, D. Wang, P. A. Heng, T. T. Wong, W. C. W. Chu, B. H. Y. Yeung and J. C. Y. Cheng, "Skull Shape Analysis in Adolescent Idiopathic Scoliosis Patients," *International Journal of Computer Assisted Radiology and Surgery, (Proceedings of CARS 2007)* Vol. 2- Supplement 1, pp. S480, June 27-30, 2007, Berlin, Germany.

[18] L. Shi, P. A. Heng, T. T. Wong, C. W. Chu, H. Y. Yeung and C. Y. Cheng, "Morphometric Analysis for Pathological Abnormality Detection in the Calvariums of Adolescent Idiopathic Scoliosis Patients," in proceedings of *MICCAI 2006*, LNCS 4190, pp. 175 – 182, Oct 1 – 6, Copenhagen, 2006.

[19] P. A. Heng, "Imaging Technologies for Orthopaedic Visualization and Simulation," *Advanced Bioimaging Technologies in Assessment of Quality of Bone and Scaffold Biomaterials*, pp. 51 – 64, edited by L. Qin, H. Genant, J. Griffith and K. S. Leung, published by *Springer Verlag*, 2007.

[20] P. A. Heng, "Research and Applications of Virtual Medicine," accepted by *27th Annual International Conference of the IEEE Engineering In Medicine and Biology Society (EMBS)*, Shanghai, September 1-4, 2005.

[21] P. A. Heng, C. Y. Cheng, T. T. Wong, W. Wu, Y. S. Xu, Y. M. Xie, Y. P. Chui, K. M. Chan and K. S. Leung, "Virtual Reality Techniques: Application to Anatomical Visualization and Orthopaedics Training", *Clinical Orthopaedics and Related Research*, Vol. 442, pp. 13-20, January 2006. (IF = 1.403)

[22] P. A. Heng, T. T. Wong, K. M. Leung, Y. P. Chui and H. Sun, "A Haptic Needle Manipulation Simulator for Chinese Acupuncture Learning and Training," *International Journal of Image and Graphics*, Vol. 6, No. 2, pp. 205 – 230, April 2006.

[23] P. A. Heng, T. T. Wong, R. Yang, Y. M. Xie, Y. P. Chui, K. S. Leung and P. C. Leung, " Intelligent Inferencing and Haptic Simulation for Chinese Acupuncture Learning and Training," *IEEE Transactions on Information Technology in Biomedicine*, Vol. 10, No. 1, pp. 28 - 41, January 2006.

[24] P. A. Heng, S. X. Zhang, Y. M. Xie, L. Shi, T. T. Wong and Y. P. Chui, "Visible Human based Virtual Medicine," *The International Journal of Virtual Reality*, Vol. 5, No. 4, pp. 13-20, December 2006.

[25] P. A. Heng, S. X. Zhang, Y. M. Xie, T. T. Wong, Y. P. Chui, and C. Y. Cheng, "Deploying Chinese Visible Human Data on Anatomical Exploration: From Western Medicine to Chinese Acupucnture," *Complex Medical Engineering*, pp. 351 – 360, edited by J. L. Wu, K. Ito, S. Tobimatsu, T. Nishida and H. Fukuyama, Springer Publisher, 2007.

[26] P. A. Heng, S. X. Zhang, Y. M. Xie, T. T. Wong, Y. P. Chui, C. Y. Cheng, "Photorealistic Virtual Anatomy Based on Chinese Visible Human Data," *Clinical Anatomy*, Vol. 19, No. 3, pp. 232-239, April 2006.

[27] P. A. Heng, Y. Xie, X. Wang, Y. P. Chui and T. T. Wong, "Virtual Acupuncture Human based on Chinese Visible Human Dataset," in *Proceedings of Medicine Meets Virtual Reality 14 Conference*, Long Beach, CA, USA, Jan 24 – 27, 2006.

[28] Q. Chen, J. Luo, P. A. Heng and D. Xia, "Fast and Active Texture Segmentation based on Orientation and Local Variance," *Journal of Visual Communication & Image Representation*, Vol. 18 (2007) 119 – 129.

[29] Q. Chen, Z. Zhou, M. Tang, P. A. Heng and D. Xia, "Shape Statistics Variational Approach for the Outer Contour Segmentation of Left Ventricle MR Images," *IEEE Transactions on Information Technology in Biomedicine*, Vol. 10, No. 3, pp. 588 – 597, July 2006.

[30] W. Y. Chan, P. W. Chow, T. W. Yew, Y. P. Chui and P. A. Heng, "An Automatic Annotation Tool for Virtual Anatomy," in Proceedings of *IEEE International Conference on Integration Technology (IEEE ICIT 2007)*, pp. 269 – 274, March 20-24, Shenzhen, China.

[31] Y. M. Xie, P. A. Heng, G. Y. Wang and T. T. Wong, "GPU-Friendly Marching Cubes for Visualizing Translucent Isosurfaces," in proceedings of *MMVR 15 (Medicine Meets Virtual Reality 2007)*, pp. 500 – 502, Feb 6 – 9, 2007, Long Beach, California, USA.

[32] Y. Qu, P. A. Heng and T. T. Wong, "Image Segmentation Using the Level Set Method," *Deformable Models: Theory and Biomaterial Applications*, pp. 95 – 122, edited by Jasjit S. Suri and Aly Farag, Springer Publisher, 2007.

[33] Y. Qu, P. A. Heng, T. T. Wong, "Semi-automatic Segmentation and Tracking of CVH Data," in *Proceedings of Medicine Meets Virtual Reality 14 Conference*, Long Beach, CA, USA, Jan 24 – 27, 2006.

Systematic Synthesis of Nano-informatics Chips by Nano-robotics Manipulation

Principal Investigator:

Professor Wen-Jung Li [1]

Co-investigator:

Professor Philip H.W. Leong [2]

International Collaborators:

Professor Steve C.H. Tung [3]
Professor Ning Xi [4]

Research Team Members:

Mr. Gary C.T. Chow, Research Assistant [2]
Ms. Carmen K.M. Fung, Research Assistant [1]
Mr. King W.C. Lai, Research Assistant [1]
Mr. Patrick G.W. Leung , Research Assistant [1]
Ms. Mandy L.Y. Sin, Research Assistant [1]
Ms. Maggie Zhang, Junior Research Assistant [1]
Ms. Jennifer W.L. Zhou, Research Assistant [1]

[1] *Dept. of Mechanical and Automation Engineering, CUHK (formerly Dept. of Automation and Computer-Aided Eng. CUHK.)*
[2] *Dept. of Computer Science and Engineering, CUHK*
[3] *Dept. of Mechanical Engineering, University of Arkansas, USA*
[4] *Dept. of Electrical and Computer Engineering, Michigan State University, USA*

Project Start Date: 1st March 2005
Completion Date: 28th February 2007

Abstract

In this project our team has developed a systematic procedure to synthesize Carbon Nanotubes (CNTs) based nano-sensing chips. We have produced a set of nano-chip synthesis rules that can be followed by researchers throughout the world, thereby, establishing SHIAE at CUHK as one of the global leaders in engineering nano devices. Our team also has demonstrated the construction of several CNT based sensors and performed detailed experimental characterization of these sensing devices. These nano-sensing chips will potentially revolutionize conventional sensing technologies spanning several technical disciplines – we have developed MEMS pressure sensors, microfluidic shear stress sensors, and alcohol vapor sensors all based on CNT sensing elements. All these sensors have the extremely important characteristics of being next generation advanced sensors: low-power (~uWatt input), high-spatial-resolution (nanometer dimensions), and low-cost to construct (less than 1$US per sensor).

1. Objectives and Significance

Almost 50 years have passed since J. Kilby and R. Noyce co-invented the Integrated Circuit (IC) technology in the winter of 1958. Without a doubt, the silicon-based IC technology has revolutionized many aspects of modern-day human life, including communications, computing, and entertainment. And, more recently, MEMS technology and advanced DNA-detection chips, both of which are derivatives of IC technology, are transforming methods that scientists and engineers used to perform physical sensing and opening up new frontiers in biotechnology, respectively. Some scientists and industrialists are concerned, however, that the well-recognized Moore's Law, which states that the number of transistors that the IC industry can place in commercial chips will double every 2 to 3 years, is rapidly reaching a limit. That is, within a decade, it may take 4 to 6 years before the number of transistors can double in a chip. There are two fundamental limits which the IC industry is facing now: 1) the smallest feature 'printable' on a chip has reached both the optical-wavelength limit and electron-beam lithography limit, and 2) the thin films deposited on the chips are reaching atomic layer thickness limit. Hence, most experts predict that a new technology has to be developed, perhaps completely different from the lithography and silicon based IC technology, to ensure the future advancement of faster and low-power-consuming micro computing processors.

On the other hand, Carbon Nanotube (CNTs), since its discovery by Iijima in 1991, has become a major candidate that is considered as a promising replacement for silicon-based electronic elements. CNTs (or nanowires in general) have been experimentally determined to have many more desirable characteristics than silicon -- such as higher yield-strength, higher thermal conductivity, and can be semi-conductive. Best of all, nanowires and CNTs can be made to have diameters of 1 to 10nm, which is at least an order of magnitude smaller than commercial IC transistors dimensions. This is the premise that IBM Zurich and Intel are both interested in developing computing chips based on CNT technology. However, due to the nano-scale dimensions of the CNTs, it is still very difficult to build and test circuit components and sensors consistently using CNTs as electronic or sensing elements. There are only 4 available techniques available today to construct CNT sensors: 1) onsite-catalytic growth (Stanford University and Samsung being the most successful groups, but controlling the conductive properties of CNTs is nearly impossible using this method), 2) SPM/AFM based nano-robotic manipulation, 3) self-assembly based on chemical bonding (no working device has been demonstrated yet using this technique), and 4) dielectrophoretic (DEP) manipulation, i.e., inducing electrical field gradient between microelectrodes to manipulate nano-entities.

Through the funding of the SHIAE, our group has achieved the following two major objectives in the past two year:

A. Development of a systematic procedure to synthesize CNT based sensors by combining dielectrophoretic, microspotting, and micro-robotics technology: This is the most important and ambitious objective of this project. Our developed method is fairly simple: use DEP manipulation coupled with micro-injection (micro-robotics and microspotting) to pre-disperse few strands (bundle) of CNTs on predetermined locations, and then, if required, use Atomic Force Microscope based manipulation to assembly the CNTs across the microelectrodes locally. Our main project goal is to formalize this approach and automate the entire process as much as possible. Our aim is to develop what Kilby and Noyce have done in defining the basic process for fabricating integrated circuit components, and what John Tanner has accomplished in creating a software tool for MEMS designers about 15 years ago. In essence, our research objective is to automate the following sequence of processes: 1) mask lay-out design using a simplified software-tool; 2) DEP manipulation of CNT bundles to localized electrodes on a chip using a custom-built micro-robotic station; 3) If required, a nano-robotic manipulation to create single-strand CNT sensing/circuit elements across electrodes on a chip. Through this funding, we have shown that the above sequence can be systematically formalized, and hence, the SHIAE is now recognized to have one of the fastest and most advanced systems in the world in producing CNT-based circuit and physical sensor arrays.

B. Synthesis and demonstration of high-density CNT based sensor arrays: Our group now has the capability to produce highly complex arrays of CNT/nanowire/nanoparticle-linked elements that can be tested as advanced electronic components and nano sensors. We have built 10x10 arrays of CNT based thermal sensors, single-column arrays of CNT based shear stress sensors integrated in microfluidic channels, and extremely low-power consuming alcohol vapor sensors and MEMS pressure sensors.

In summary, we have developed a systematic process to produce CNT based devices reliably and repeatably by combining DEP manipulation, micro-robotics, and micro-injection – the resulting process is a foundation technology that will advance nano-manufacturing in the coming decades. Through the use of this process, we have demonstrated nanometere-scale, extreme low-power-consuming, and low-cost novel sensors that are poised to replaced existing sensors across a wide spectrum of engineering and scientific applications.

2. Research Methodology

As stated earlier, our most ambitious objective in this project is to develop a systematic approach to fabricate CNT devices repeatably and reliably. If this can be accomplished, then SHIAE/CUHK will be recognized worldwide as a key developer of advanced nano-manufacturing technologies. In addition, the development of this foundation technology for nano-manufacturing will enable our team to continue to construct and characterize many novel nano-sensing devices in the coming years. We will describe below the major challenges in achieving this goal and our solutions to overcome these challenges.

A computer-controlled CNT microspotting system was developed to perform the dielectrophoretic CNT manipulation automatically. This system allows the CNT/ethanol solution to be ejected between each pairs of microelectrodes precisely. The volume of CNT/ethanol droplet can also be well controlled, resulting in a high yield and high precision batch assembly method for CNT based devices. The system development included dielectrophoretic CNT manipulation, fluid ejection through fabricated micro probes, and position control of the micromanipulator placement between the microelectrodes on the substrate. In order to eject CNT/ethanol solution spots precisely to the microelectrodes, several key technical issues had to be resolved.

A. Fabrication of Nanometric Probes

The micro probe tip profile directly affects the size of the micro spot. Smaller probe tip diameters allow the system to eject smaller size droplets. Capillary probes were fabricated by our novel chemical etching process, which employed glass tubing as a sacrificial barrier, to control and sharpen the probe into different tip profiles. The original capillary tubing (TSP002150 with inner diameter of 2 microns and outer diameter of 126 microns, which was the smallest size provided by Polymicro Technologies) can be sharpen into capillary probe (with same inner diameter and smallest outer diameter of 6.5 microns).

B. Microinjection Experiment

Experiments were performed to understand the microfluid injection processes. The system consists of five major components: capillary probes, a syringe pump, a programmable micromanipulator, a CCD video microscope system and a computer. The whole system is installed on a vibration isolation table. Before the injection experiment, a silicon substrate was placed on the stage. A fabricated capillary probe was mounted on a programmable X-Y-Z micromanipulator (MP285, Shutter Instrument Company), which was used to move the probe to the desired position of the substrate. A syringe pump (V6 syringe drive modules, Kloehn Limited) was connected to the capillary probe and was used to drive the probe to eject the fluid droplets to the substrate. It was successfully demonstrated the ultra small fluid droplet (~40 um) could be ejected by using our fabricated probe (~20 um outer diameter). It was found that the adhesion force between the outlet of the capillary probe and the fluid droplet was very large during the experiment, i.e. fluid cannot be ejected. Fluid typically comes out from the outlet of the capillary probe, but sticks on the capillary probe side wall until the droplet becomes very large. The effect of adhesion force can be eliminated by different methods such as reducing the gap distance between the probe tip and the substrate, i.e., making the probe tip touching the substrate (droplet contact method), or increasing injection pressure.

C. CNT/ethanol Solution Spotting Process on Microelectrodes

After understanding the dynamics of the injection process, we decided to use the droplet contact method to eject the CNT/ethanol solution onto a substrate. The experiment was conducted by using our fabricated probes (10 um inner diameter, 40 um outer diameter). The CNT/ethanol droplet sizes achieved were approximately 50 um. Since the microchips with arrays of microelectrodes were fabricated by a standard microlithographic technique, a CIF-mask computer file is required and pre-designed. A custom CNT layer, which can be defined in the CIF-mask computer file, is added to locate different positions of microelectrodes. The CIF-mask computer file is imported to our automated CNT microspotting system. All positions of microelectrodes is obtained and stored in an X-Y coordinate system array. The upper-left microelectrode is defined as the initial reference position, the next position is obtained by finding the nearest neighbor from remaining microelectrode pairs (using a well-know sorting principle). The probe is attached to the programmable X-Y-Z micromanipulator (MP285, Shutter Instrument Company), and a CCD camera is connected with the microscope (Micromanipulator P7000) to locate the initial position of the array of microelectrodes through the computer system. The probe tip is required to align to an initial microelectrode. After the initial position was calibrated, the spotting process could begin. The probe tip was moved to positions which were above each pairs of microelectrode, then it was moved downwards to the substrate, and the syringe pump ejected the CNT/ethanol solution from the probe to the substrate.

D. CNT Manipulation using DEP

An AC field was applied to the microelectrodes before dropping the CNT/ethanol solution. After dropping the CNT/ethanol solution on the pair of microelectrodes, the ethanol evaporated away (after a few seconds) leaving the MWNTs to reside between the gaps of the microelectrodes. We observed from experimental results that bundled MWNTs were attracted towards the Au microelectrodes under non-uniform electric field and connected across the microelectrodes. In order to confirm the linkage of bundled CNTs between each pairs of microelectrodes, the room temperature resistance corresponding to each pair of microelectrodes was measured. The CNT connection process was deemed successful between 2 microelectrodes when the room temperature resistance measured became several kΩ to several hundred kΩ. During the dielectrophoresis process to form MWNT bundles across microelectrodes, the MWNTs were randomly connected between microelectrodes. Therefore, it is logical that different MWNT samples exhibited different conductivities.

To validate the consistency of the batch assembly of CNT devices, repeated experiments for different chips were performed. From the experimental results, we have observed that the range of the room temperature resistances for CNTs is from several kΩ to several hundred kΩ as stated earlier. We have experimentally found that the success rate for different sensor chips is consistent with overall success rate, which is equal or greater than 80%. The success rate is defined as the ratio of the number of successful CNT-connected microelectrodes to the total number of microelectrodes on a substrate.

3. Results Achieved

We have accomplished the following major technical results at the conclusion of this project:

a) Development of an infrastructure consisting of both hardware and software to demonstrate automatic synthesis of CNT based nano-informatics chips. The system combines micro-robotics, micro-injection, and DEP manipulation to build CNT sensors efficiently and reproducibly.

b) Successful built and characterized various prototype CNT based sensing chips for applications such as aerodynamic/microfluidic and biological/chemical measurements.

The following list some of the specific technical milestones achieved using the funding from SHIAE:

- **Automating the CNT Sensor Fabrication Process:** An automated Carbon Nanotube (CNT) microspotting system was developed for rapid and batch assembly of bulk multi-walled carbon nanotubes (MWNT) based MEMS sensors. By using dielectrophoretic and

microspotting technique, MWNT bundles were successfully and repeatably repeatedly manipulated between an array of micro-fabricated electrodes. Preliminary experimental results showed that over 80% of CNT functional devices can be assembled successfully using our technique, which is considered to be a good yield for nanodevices manufacturing. This system was demonstrated to be capable of producing a 10x10 CNT sensor array and requires only ~1sec to form the CNT sensing element across each sensor.

- **CNT Based MEMS Pressure Sensor:** We have developed and characterized a novel MEMS sensor using Carbon nanotubes (CNTs) as sensing elements. Using a MEMS-compatible process and the dielectrophoretic (DEP) nanoassembly of CNTs, we have successfully integrated bundled strands of CNT sensing elements on arrays of Polymethylmethacrylate (PMMA) diaphragms. With a thin film of Parylene C covering the CNT bundles and their contacts to electrodes, the CNT bundles show consistent and repeatable piezoresistivity from multiple electromechanical measurements. Using low noise data acquisition techniques and annealing treatments of the CNTs, the fluctuations of the signal are greatly reduced. Based on these experimental evidences, we propose that embedded in Parylene-C thin film, CNT bundles is a novel material for fabricating micro pressure sensors on polymer substrates – which may serve as ultra-low-power (μW) and low noise sensors when bio-compatibility and low-cost applications are required.

- **Ultra-Low-Power Consumption Alcohal Vapor Sensor:** Our group has fabricated alcohol vapor sensors by forming bundles of chemically functionalized multi-walled carbon nanotubes (f-CNTs) across Au electrodes on SiO2/Si substrates using an AC electrophoretic technique, and showed that these sensors only require ~ 0.01-1 μW power to activate, which is lower than the power required for most commercially available alcohol sensors by more than 4 orders of magnitude. The multi-walled carbon nanotubes (MWCNTs) used in these sensors were chemically functionalized with the COOH groups by oxidation. We found that the sensors are selective with respect to flow from air, water vapor, and alcohol vapor. The sensor response is linear for alcohol vapor concentrations from 1 to 21 ppm with a detection limit of 0.9 ppm. The transient response of these sensors is experimentally shown to be ~1 s and the variation of the responses at each concentration is within 10 % for all of the tested sensors. The sensors could also easily be reset to their initial states by annealing the f-CNTs sensing elements at a current of 100-200 μA within ~100-200 s. We demonstrated that the response of the sensors can be increased by 1 order of magnitude after adding the functional group COOH onto the nanotubes, i.e., from ~0.9 % of a bare MWCNTs sensor to ~9.6 % of an f-CNTs sensor with a dose of 21 ppm alcohol vapor.

- **Ultra-Low-Power Consumption Microfludic Shear Stress Sensor:** We have developed CNT sensors for gas-flow shear stress measurement inside a Polymethylmethacrylate (PMMA) microchannel. An array of sensors is fabricated by using dielectrophoretic (DEP) technique to manipulate bundled single-walled carbon nanotubes (SWNTs) across the gold microelectrodes on a PMMA substrate. The sensors are then integrated in a PMMA microchannel, which is fabricated by SU-8 molding/hot-embossing technique. Since the sensors detect gas-flow by thermal transfer principle, we have first examined the I-V characteristics of the sensors and confirmed that self-heating effect occurs when the input voltage is above ~1V. We then performed the flow sensing experiment on the sensors using constant temperature (CT) configuration. The voltage output of the sensors increases with the increasing flow rate in the microchannel. We also found that the power of the sensors has a linear relation with 1/3 power of the shear stress. Moreover, measurements of sensors with different overheat ratios were compared and results showed that sensor is more sensitive to the flow with a higher overheat ratio.

Based on the fundamental results obtained from this project, we were able to apply successfuly for various funding from the Hong Kong and Chinese governments successfully. These new fundings will allow our group to sustain our high level of academic research while looking for an opportunity to commercialize the technologies developed in the project. Grants received so far include the following:

- Wen J. Li, Principal Investigator of the Chinese National 863 Plan project on "面向微納製造的DEP自動化裝配技術研究" (Automated Micro/Nano Manufacturing Based on DEP Assembly), (Project Ref. No.: 2007AA04Z317), RMB$850,000, 2007-2009.

- Wen J. Li, , Principal Investigator of the Hong Kong RGC Grant on "Fabrication of CMOS-Integrated Carbon Nanotube Sensors by DEP and Nano-Spotting Technologies", HK$1,204,500, (Project code: 413906), 2006 to 2009. PI: Wen J. Li.

- Wen J. Li, Principal Investigator of the Nation Natural Science Foundation of China (NSFC) Grant on "Theoretical Investigation of DEP-based Manipulation of Micro and Nano Particles", RMB$270,000; project number: 60675060, 2007-2009 ("基於介電泳的微/納米粒子並行操縱基礎理論方法研究" , 國家自然科學基金;項目號: 60675060).

- Wen J. Li, Project Coordinator/PI of the Innovation Technology Fund Project on "Development of Ultra-Low-Power Alcohol Vapor Sensors Based on Functionalized CNT Sensing Elements (ITS/027/06), HK$1,021,700, 01/09/2006 to 31/08/2007.

Through the support of SHIAE, we have trained the following students and researchers in the last few years. We are very excited to report that all of them have gone on to further adavance their intellectual persuits in the field of Nanotechnology worldwide. Listed below are they current positions after leaving our research group.

- Maggie ZHANG, graduate student, University of California, Berkeley, USA
- Patrick LEUNG, graduate student, Cambridge University, UK
- Mandy SIN, PhD student, University of Arizona, USA
- Gary CHOW, PhD student, Imperial College, UK
- King W. C. Lai, Postdoctoral Fellow, Michigan State University, USA
- Carmen K. M. Fung, Postdoctoral Fellow, Michigan State University, USA
- Jennifer W. L. Zhou, Associate Professor, Huazhong Univ. of Science and Technology, China

Through the above accomplishments, we have promoted the SHIAE at the international arena for its commitment to achieve the highest level of engineering and scientific research. We have also published our research results in some of the top engineering and scientific conferences and journals related to MEMS and Nanotechnology. A list of our publications are listed in the next section. Please also note that several journal papers are still under preparation, and if they are accepted, the SHIAE's support will also be acknowledged.

4. Publications and Awards

The fund provided by this project enabled our team to produce some world-leading results in carbon nanotubes manipulation and nano-sensing applications, and our research results were published a the following technical publications, with acknowledgement to SHIAE's support.

[1] Minglin Li, Yanli Qu, Zaili Dong, Yuechao Wang, and Wen J. Li, "Limitations of Au Particle Nano-Assembly using Dielectrophoretic Force-A Parametric Experimental and Theoretical Study", accepted, *IEEE Transactions on Nanotechnology (Letter)*, November 2007.

[2] Leung, Gong Wai; Lau, Fong Ting; Leung, Siu Ling; Li, Wen J.; "Formation of Au Colloidal Crystals for Optical Sensing by DEP-Based Nano-Assembly", *2nd IEEE Int. Conf. on Nano/Micro Engineered and Molecular Systems, 2007, IEEE-NEMS 2007*, Jan. 2007 Pp. 922 – 926.

[3] Mandy LY Sin, Gary CT Chow, Gary MK Wong, Wen J Li, Philip HW Leong, Ka Wai Wong, "Ultra-Low-Power Alcohol Vapor Sensors using Chemically Functionalized Multi-Walled Carbon Nanotubes," *IEEE Transactions on Nanotechnology*, May 01, 2007.

[4] Mandy L. Y. Sin, Gary C. T. Chow, Carmen K. M. Fung, Wen J. Li, Philip Leong, K. W. Wong, and Terry Lee , "Ultra-Low-Power Alcohol Vapor Sensors Based on Multi-Walled Carbon Nanotube ," *1st IEEE Int. Conf. on Nano/Micro Engineered and Molecular Systems, IEEE-NEMS 2006* , January 2006.

[5] Steve Tung, Husein Rokadia, and Wen J. Li, "A Micro Shear Stress Sensor Based on Laterally Aligned Carbon Nanotubes," *Sensors and Actuators A: Physical*, June 15, 2006.

[6] Mandy L. Y. Sin, Gary C. T. Chow, M. K. Wong , Wen J. Li, Philip Leong, K. W. Wong, and Terry Lee, "Chemically Functionalized Multi-Walled Carbon Nanotube Sensors for Ultra-Low-Power Alcohol Vapor Detection ," *IEEE-NANO 2006*, July 16, 2006.

Prof. Li was honored to give the following invited speeches worldwide due the Nanotechnology community's interest in our group's CNT sensor work, which was partially supported by this SHIAE project. The following Invited Presentations by Prof. Wen J. Li have acknowledged the support of SHIAE.

- Wen J. Li, **Plenary Speaker**, The 24th Sensor Symp. on Sensors, Micromachine and Applied Systems, October 16-17, 2007, Tokyo, Japan. (http://www2.iee.or.jp/~smas/Sensor_Sympo/24)

- Wen J. Li, **Special Invited Speaker**, "Towards Fabrication of CNT-based Large-Scale-Integrated (LSI) Sensor Arrays", The 2nd International Meeting on Microsensors and Microsystems, Tainan, Taiwan, 15-18 January 2006. (http://imu2.ncku.edu.tw/)

- Wen J. Li, **Keynote Speaker**, "Reversible and Ultra-Low-Power Alcohol Vapor Sensors based on Functionalized CNT Elements ", The 2nd International Symposium on Micro and Nano Technology, Hsinchu, Taiwan, March 29-31, 2006. (http://www.mems.nthu.edu.tw/ismnt-2/)

- Wen J. Li, **Special Invited Speaker**, "Rapid Fabrication of Functional CNT Sensor Arrays using Micro- spotting and DEP Technologies", IEEE Nanotechnology Materials and Devices Conference, October 22-25, 2006, Gyeongju, Korea. (http://www.ieee-nmdc.org)

- Wen J. Li, **Plenary Speaker**, "Micro-Robotics and Nano-Fluidics: Critical Technologies to Building LSI Nanosystems", IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2005), Edmonton, Alberta, Canada, Aug 04, 2005 (www.iros2005.org).

- Wen J. Li, **Special Invited Speaker**, "Fabrication of Large CNT Sensing Arrays by Pico-Litre Fluidic Spotting and Dielectrophoretic Manipulation", The International Conference on Bio-Nano-Informatics Fusion, Los Angeles, CA, USA, July 21, 2005 (www.bnifusion.org).

Study on RF Radiation Effect and Signal Efficiency of Wireless Medical Devices

Principal Investigator:

Professor Max Q.-H. Meng [1]

Co-investigators:
Professor Yuan-Ting Zhang [1]
Professor Ke-Li Wu [1]
Professor Justin Wu [2]

Research Team Members:

Dr. Lisheng Xu, Postdoctoral Fellow [1]
Dr. Xiaona Wang, Research Assistant [1]
Dr. Rong Liu, Research Assistant [1]
Mr. Hongliang Ren, PhD student [1]

[1] *Dept. of Electronic Engineering, CUHK*
[2] *Dept. of Medicine and Therapeutics, CUHK*

Project Start Date: 1st May 2006

Abstract

In this project, we have systematically studied the radiation effects and communication link performance of ingested wireless devices. We analyzed the radiation characteristics of implantable and ingestible exposure sources in 5-year old children, 10-year old children and adults via numerical simulations and found that the radiation efficiency changed from -53.39 dB to -116.4 dB. We recommend that different transmission power levels should be used for children and adults, respectively. We also studied the variation of radiation efficiency in the bi-directional communication of the ingestible devices in relation to the distance between the antenna and a homogenous human trunk model. The simulation results indicate that the distance has little influence on the communication performance, although the peak of specific absorption rate (SAR), the maximum of the 1-g and 10-g averaged SAR values, and the port impedances all vary significantly in relation to the distance. This is further validated via simulation studies of a realistic human body model. Through the study of the radiation characteristics when the data was transmitted from the ingested device to the outside receiver, we investigated if the frequency of 2.4GHz is safe among the most efficient frequencies for medical clinical applications. In our study, we found that the influence of human legs and head on the maximum SARs is quite negligible. We thus used the human body model without the legs and head as a reference for the design of human body trunk phantom for testing the compliance with the safety of body-worn, implantable and ingestible wireless medical devices. Finally, we studied the applicability of using the above homogeneous tissue phantoms in place of a complex multi-tissue human model and then used the human body trunk phantom with human body tissue equivalent liquid for the study of the determination of peak 1-g and 10-g SAR needed for compliance with various safety standards.

1. Objectives and Significance

The radiation characteristics of implantable and ingestible wireless devices are scarcely studied until recently. Our research in this project provides a better guideline for compliance checking and design of various wireless medical devices, in addition to our work on the optimization of communication link performance in relation to antenna and medical sensory devices. Furthermore, this project proposed and used a conservative homogeneous body trunk model for efficient estimation of the radiation effects and link performance so that more accurate compliance tests and design on the body-worn and ingestible wireless devices can be achieved.

2. Research Methodology and Results Achieved So Far

2.1. Effect of trunk size on electromagnetic radiation from a source inside human body

There have been many investigations on biologic effects in the heads of children and adults using various exposure scenarios and models. Martinez-Burdalo et al. analyzed the specific absorption rate (SAR) deposition in different sizes of human head exposed to the electromagnetic radiation from a mobile phone. However, investigations on the human body trunk are rarely found. For assessment of compliance with the safety guidelines, the radiation effects and efficiency of ingested wireless medical devices inside human body trunk models of five-year-old children, ten-year-old children and an adult are studied by using developed homogenous models and a half-wavelength dipole at the frequency of 2.4 GHz.

Numerous simulations were conducted using nine body trunk models, including three adults, three ten-year-old children, and three five-year-old children. The dimensions of the three models at each age group represent the statistical results of small, medium and large sizes of their age groups, respectively. The finite-difference time-domain (FDTD) method is used for analyzing the radiation effects and efficiency. In the simulation studies, the input power is adjusted from the computed value to the specified 1W. The simulation results indicate that the radiation power was attenuated very rapidly and almost all the radiation power was dissipated in the tissues of the body trunk models. However, the radiation efficiency and system efficiency of the models vary quite significantly as presented in Table I.

Due to the short penetration depth, the SAR distributions of these models are almost same. SAR limits are based on whole-body exposure levels of 0.4 W/kg for RF workers and 0.08 W/kg for the general population. The larger the dose of exposure is, the larger the effect is. Actually, human body absorbed almost all of the radiated power. In order to reduce the radiation harm of wireless device and to meet the safety standards for radiation exposure, we recommend using different power levels of the wireless ingestible or implantable devices for children and adults, considering the tradeoff between radiation effect and efficiency.

		Volume ($X \times Y \times Z$) in cm^3	Radiation Efficiency in dB	System Efficiency in dB
5-year old	S	15×10×8	-53.39	-55.31
	M	20×15×10	-61.36	-63.29
	L	20×17×12	-68.89	-70.81
10-year old	S	25×18×10	-61.26	-63.19
	M	30×20×10	-61.22	-63.15
	L	30×22×15	-80.17	-82.09
Adult	S	45×20×15	-80.15	-82.07
	M	50×30×20	-99.03	-100.94
	L	60×35×25	-116.43	-118.35

Table I: Radiation and system efficiencies of the 3 models of adult, 10-year-old and 5-year-old children.

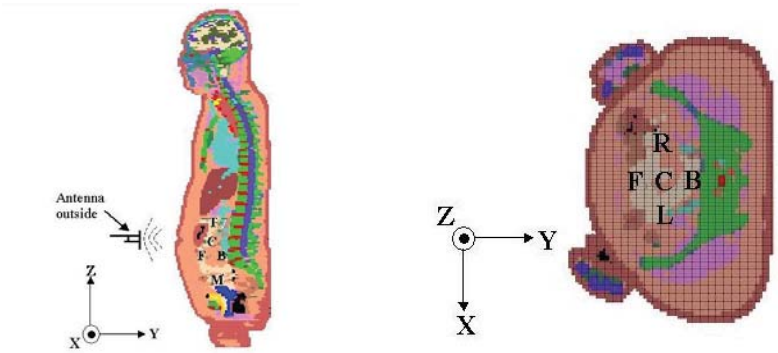
2.2. Variation of radiation effects with distance between electromagnetic source and trunk model

The concerns regarding the radiation effects and link performance of wireless devices have emerged again due to the increasing application of implanted and ingested bi-directional wireless devices. In our study, the FDTD method is applied to analyze the variation of biological effects and signal efficiency with distance between the electromagnetic source and the human body models. Considering the tradeoff between radiation effects and the attenuation of wireless signals, we choose the frequency of 900MHz to transmit data and the control commands. FDTD simulations are carried out for the 900MHz systems with a half-wavelength dipole antenna. The distance is changed from 25 mm to 1 mm within the range of $\lambda/2$. The SAR distributions and electric fields for various vertical and horizontal slices of the human trunk are calculated. The computational results are interpreted in terms of international safety guidelines for human health, showing that distance has little influence on signal efficiency, with 1 Watt input power, the peak SAR, the maximums of 1-g and 10-g averaged SAR values are 6.30 W/kg, 4.55 W/kg, and 3.18 W/kg, respectively. The ingested wireless medical devices, with transmission power of less than a few mWs, will satisfy the safety guidelines. Considering the radiation efficiency of wireless devices and safety guidelines for radiation exposure, we recommend that the antenna outside the human body should be kept at a suitable distance (i.e., 12 mm) from the human body. The simulation studies of a realistic female human body model validate this conclusion.

2.3. Electromagnetic radiation characteristics from ingested sources inside the human intestine

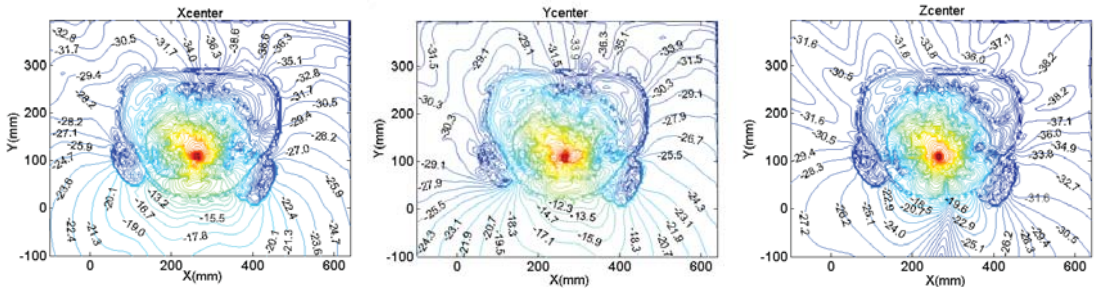
The widespread use of ingested wireless medical devices during the past few years has attracted a great deal of social and scientific concerns about the possible harmful effects of RF exposure. In the commercially available capsule endoscopes, the 300-500 MHz is the most used frequency band. However, higher resolution images of the gastrointestinal (GI) tract with faster transmissions are demanded for more accurate diagnosis. The bandwidths of the compact antennas working at low frequency and the signal efficiency are both quite low. Among the frequency ranges of the industrial, scientific, and medical (ISM) applications, the 2.4GHz, widely used in commercial applications, can be a good choice. Since more energy will be absorbed by human tissues for higher frequency signals, many questions remain to be answered, such as the safety issues and the influence of the position and orientation of the RF source on the radiation characteristics. We studied the radiation effects and wireless signal efficiency of an ingested source in a realistic human body model at the frequency of 2.4GHz in 21 scenarios in order to investigate whether 2.4GHz are suitable for use for ingested devices. We assume that the signals transmitted from the helix antenna at seven locations with three orientations are received by an antenna placed at the center position above the abdomen as shown in Fig. 1. The FDTD method is applied to analyze the electromagnetic interaction with the human body and the Penns bio-heat equation is applied to calculate the temperature rise due to the SAR deposition. The port impedances and the radiation efficiency from the simulation study vary greatly with the location and orientation changes. The impedances along the X and Y orientations are similar, while the impedances of the Z orientation are higher. When the ingested device is positioned at B, the power reflected by the tissues is the most while at F, less power is reflected by human tissues than at the other six positions. The simulation results indicate that most of the radiation power was dissipated in the tissues. When the ingested source is oriented along X and located at C, the temperature rise reaches maximal at 0.029℃ with larger SAR value.

Fig. 2 illustrates the electric fields of the XY planar slice where the outside receiving antenna was located, including the electric fields inside and outside the human body, where the ingested antenna is positioned at the seven positions with X, Y and Z orientations respectively. In these figures, we can find that the near fields of the antenna at the X, Y and Z orientations are differently distributed. The electric fields in the anterior part of the human body are bigger than those in the posterior. All the distributions of the electric fields outside human body are irregular and asymmetrical due to the heterogeneous tissues. When the ingested antenna is located in position F, the near fields are larger than those of the other six positions. When the ingested antenna is in position B, the near fields are less than those in the other six positions. With 21 simulation scenarios, we found that 2.4GHz can be used for efficient transmission of video signal in wireless capsule endoscope, the maximum temperature rise is 0.078℃ when the delivered power is 100 mWs that is more than what we needed for the wireless capsule endoscope to work efficiently, and the energy in the battery of a commercial wireless capsule endoscope will last more than 8 hours. In conclusion, 2.4GHz can be safely and efficiently used to transmit the video and command signals in devices such as wireless capsule endoscopes.



(a) (b)

Fig. 1. Simulation setup (a) Sagittal slice of the human model (b) Transverse slice of the human model

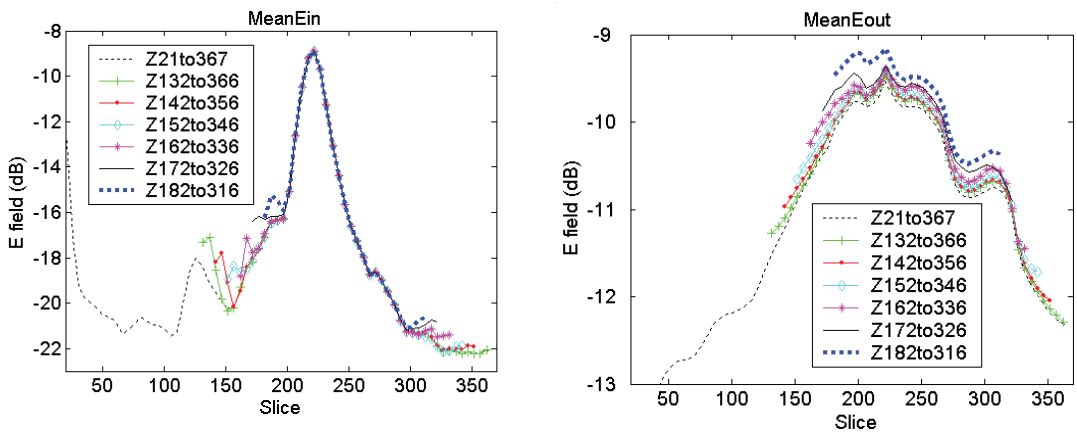


(a) (b) (c)

Fig. 2. Near fields of the XY planar slice where the receiving antenna is located exposed to an ingestible 2.4GHz source at position C: (a) X orientation; (b) Y orientation; (c) Z orientation.

2.4. Influence of head and legs on the realistic human body model simulation

Current SAR test standards and system are proposed for body-worn, implantable, and ingestible wireless medical devices. In this project we studied the influence of the head and legs on the signal efficiency and found that both head and legs can influence the distributions of electric fields, SAR, and temperature rise, but rather insignificantly. Fig. 3 shows that the electric fields of the human body horizontal slices are nearly superposed. The influence on the electric field in human body is less than the external electric field illustrated in Fig. 3. The simplified human body model without the head and the legs can be used to analyze the radiation characteristics of the RF source in the abdomen and can be used as a reference for the human body trunk phantom for testing the safety compliance of body-worn, implantable, and ingestible medical devices.



(a) (b)

Fig. 3. Distribution of the Y-axis electric field of the human body models: (a) inside; (b) outside.

2.5. The suitability of using homogeneous phantom for compliance testing

Recent research reported that layered tissues can lead to increased SAR in comparison to the homogeneous phantom. Kivekas et al claims that the SAR distribution was related to the real part of the tissue permittivity. The suitability of homogeneous head phantoms has been studied for the 900MHz frequency band. Andreas Christ et al. studied the dependence of electromagnetic energy absorption on human body tissue composition and found a significant increase of 2.2-4.7dB of the SAR peak value in comparison to the results assessed with homogeneous standard liquids. Our simulation studies for assessing the suitability of using homogeneous phantom for RF compliance testing of wireless medical devices used simplified homogeneous phantom and anatomical human body model. The simulation results indicate that the distance has little influence on the radiation power, while the peak of SAR, the maximums of 1-g and 10-g averaged SAR values, and port impedances vary greatly with the distance. Experimental phantoms using human body tissue equivalent liquid can be used for the compliance testing with various safety standards.

3. Publications and Awards

[1] L.S. Xu, Max Q.-H. Meng and H.L. Ren, "Variation of radiation effects and signal efficiency with distance between electromagnetic source and trunk model," in *Proc. of the 29th International Conference of the IEEE Engineering in Medicine and Biology Society*, August 22-26, 2007, pp. 1184-1187.

[2] L.S. Xu, Max Q.-H. Meng and H.L. Ren, "Effect of subject size on electromagnetic radiation from source in human body following 2450MHz radio frequency exposure," in *Proc. of the 2007 IEEE International Conference on Integration Technology*, March 20-23, 2007, pp. 326-329.

[3] Xu, L.S., Max Q.-H. Meng and H.L. Ren, "Electromagnetic radiation from ingested sources in the human intestine at the frequency of 2.4GHz," Accepted for presentation at the *23rd Progress in Electromagnetics Research Symposium*, Hangzhou, China, March 24-28, 2008.

[4] Xu, L.S., Max Q.-H. Meng and H.L. Ren, "Radiation characteristics of ingested sources in the human intestine at the frequency of 2.4GHz," Submitted to *IEE Electronics Letters*.

[5] Xu, L.S., Max Q.-H. Meng and H.L. Ren, "Applicability of homogeneous human trunk phantom in estimating the radiation characteristics of body-worn devices," Accepted by *International Journal of Information Acquisition*.

[6] Xu, L.S., Max Q.-H. Meng and H.L. Ren, "Radiation characteristics of 2.4GHz ingestible wireless devices in human intestine," Submitted to *IEEE Transactions on Biomedical Engineering.*

Photonic Biosensor Micro-arrays for Molecular Diagnostics Applications

Principal Investigator:

Professor Aaron H.P. Ho [1]

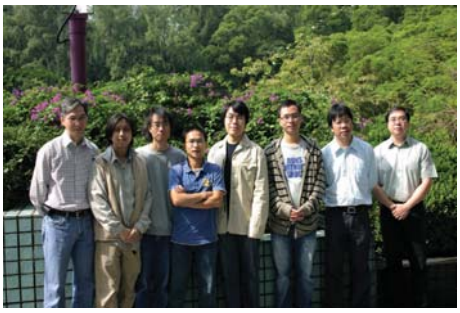
Co-investigator:

Professor S.K. Kong [2]

Research Team Members:

Mr. Shu Yuen Wu, Research Assistant [1]
Mr. Kwong Chun Lo, Research Assistant [1]
Mr. Wing Wai Wong, Research Assistant [1]
Mr. Chi Lok Wong, Research Assistant [1]
Ms. Rebecca K.Y. Lee, Research Assistant [1]
Mr. W. Yuan, Research student [1]
Mr. Arthur T.T. Yu, Research student [1]

[1] *Dept. of Electronic Engineering, CUHK*
[2] *Dept. of Biochemistry, CUHK*



Project Start Date: 1st May 2006

Abstract

This project aims to develop a novel surface plasmon resonance (SPR) photonic biosensor array platform that offers parallel detection of a range of biomolecular species without the use of any labeling procedures, thus drastically reducing the cost of performing such tests in a clinical environment. After having experimentally demonstrated the feasibility of this idea in the first part of this project, we have been mainly concentrating our work on achieving high detection resolution in our biosensor arrays through improvement of signal analysis of SPR images. We have also been working on developing DNA detection capability. Both objectives have been achieved and we are now in the process of using the system for more practical biomolecular detection applications. Within our work, we have also explored a new cross-polarization detection approach for further improvement of sensing resolution.

1. Objectives and Significance

The objectives of this project include implementation of a SPR photonic biosensor platform in which micro-arrays of sensor sites can provide parallel identification for a range of molecular species that are of relevance to diagnostics applications, and development of novel SPR-based techniques as an enabling technology for other photon-based or molecular techniques.

The significance of this project is in field of biochips for diagnostics applications. Common biochips containing arrays of protein or DNA markers are being used as diagnostic devices for screening a variety of health-related molecular properties. The detection strategy has been based on chemiluminescence, i.e. emission of light upon detecting the presence of the target protein. However, this technique requires a proprietary process of inserting florescent

label tags to the protein molecules. The reagents are extremely costly. In addition, the amount of chemiluminescence rapidly deteriorates as the detection process continues, thus making it difficulty for this technique to provide quantitative information. In this project we propose to use the surface plasmon resonance (SPR) technique for detecting the target biomolecules without having to use any florescence. The advantages are two-fold. First, removing the need of chemiluminescence will bypass any need of expensive proprietary biomolecule species. Second, SPR is a real-time quantitative technique that provides accurate measurement on the concentration of the target biomolecules. Despite these important attributes, chemiluminescence has been generally preferred by the market because it offers relatively better detection sensitivity, which is a more important factor with regard to diagnostics products.

2. Research Methodology

The project involves the implementation of a SPR imaging system that can perform SPR phase analysis on a 2-dimensional scale, thus achieving arrayed biosensing detection. The proposed system is based on a new SPR photonic sensor design previously reported by the PI. The implementation steps include first demonstrating the feasibility of the SPR imaging technique for arrayed biosensing and then using the system for detecting real biomolecular species such as DNA fragments associated with certain health conditions.

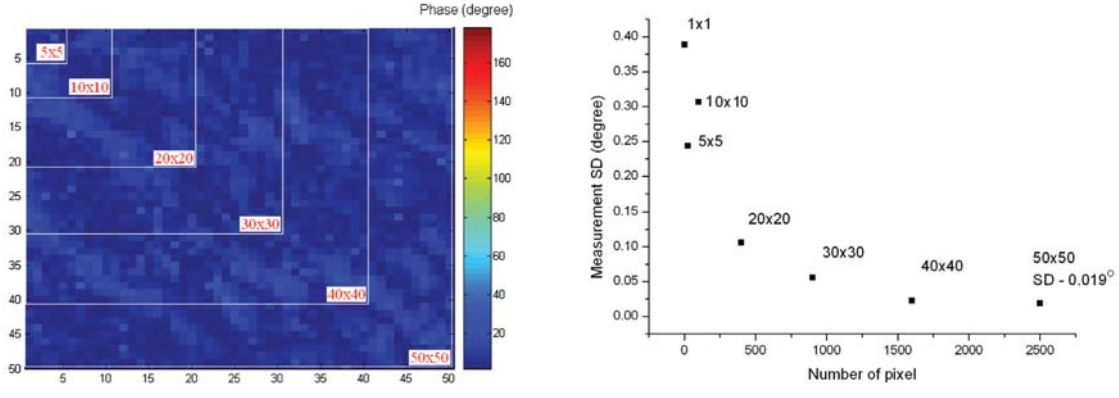
3. Project Results

3.1. Imaging SPR with improved sensing resolution

As this project's main aim is to detect practical biomolecular species, among which the DNA molecules are of the highest importance. However, because of their low molecular weight, non-labeled detection of DNA using SPR has been a well-known major challenge within the SPR instrumentation community. To this end, our effort has been concentrated on developing image analysis algorithms for obtaining SPR phase distribution on a 2-dimensional scale. We have also tested the importance of using different imaging devices.

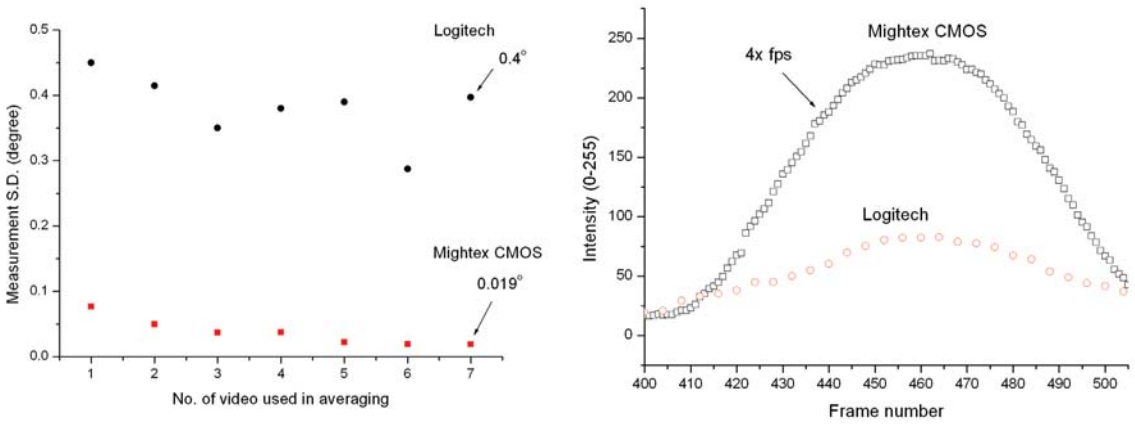
As shown in Figure 1, we have found that indeed when we use a large number of pixels within the SPR phase image for calculating each phase data point, the measurement standard deviation (SD) decreases accordingly. In our case, a best result of SD = 0.019 degree has been achieved by using 50×50 pixels, which corresponds to a sensing resolution of ~10^{-7} refractive index unit (RIU). We believe that such a resolution should be sufficient for detecting DNA binding activities.

Figure 2 shows the effect of phase imaging frame rate on the measurement accuracy of the SPR phase. Higher frame rate will substantially improve the SPR phase accuracy. This is due to the fact that a more smooth sine wave extrapolation can be obtained if we use more dense data points within the intensity waveform generated during phase-stepping.



(a) (b)

Figure 1: Measurement of SPR phase standard deviation (SD) from a 2-dimensional sensor surface in contact with water over a period of 40 minutes. The SPR phase SD decreases with increasing number of pixel included in the SPR phase extraction region. If we use a maximum of 50×50 pixels for each SPR phase data point, the measurement SD becomes 0.019 degree. The CCD device is Mightex CMOS.

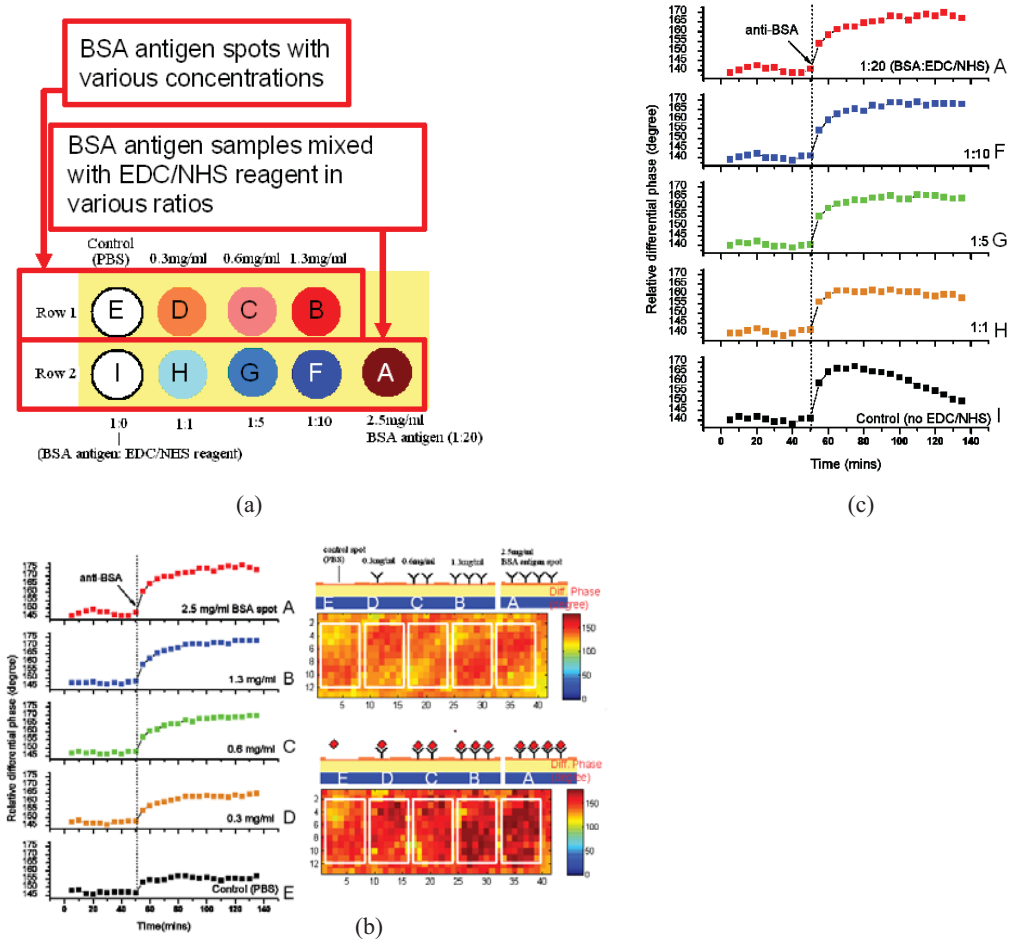


(a) (b)

Figure 2: SPR phase SD comparison between Mightex (120fps) and Logitech (30fps) cameras. Mightex provides >20 times improvement in SD from 0.4 to 0.019 degree.

3.2. Antigen-antibody detection using SPR sensor arrays

We have completed the hardware and software development for real-time SPR phase monitoring of biosensor arrays. The system subsequently used for arrayed biosensing of BSA antigen-antibody binding reactions with various concentrations. Results of a typical arrayed biosensing experiment are depicted in Figure 3.



(a) (c)

(b)

Figure 3:
(a) 9-element SPR biosensor array for detection BSA antigen-antibody binding reaction with various concentrations and EDC/NHS ratios;
(b) real-time SPR phase measurement plot and corresponding SPR phase map showing gradual increase of SPR phase caused by binding reaction, larger phase change with higher concentration of BSA antigen;
(c) effect of using various EDC/NHS ratio showing that a ratio of at least 1:10 should be used in order to obtain reliable results.

3.3. Detection of Herpes virus DNA

While the SPR phase imaging work is ongoing, we have conducted parallel experiments using a single-beam liquid crystal modulator (LCM) based SPR phase system that has proven record of providing high sensing resolution to demonstrate the possibility of our SPR phase technique is capable for label-free DNA detection in light of the requirement of extremely high resolution. The detection procedures first involve preparing the gold sensor surface for DNA detection using immobilized Avidin and then followed by the steps listed in Table 1. The corresponding SPR phase response curve is shown in Figure 4.

Step	Description
1	Inject 75ul of 4μM Probe 1 DNA [Probe 1 contains biotin which binds the Avidin on gold surface and form immobilized DNA probe layer]
2	Inject Taq buffer to rinse out excess Probe 1 DNA
3	Inject 75ul of 4μM Probe 2 DNA to sensor surface [verify the signal of Probe 2 DNA when target DNA is not present]
4	Inject Taq buffer to rinse out non-attached Probe 2 DNA
5	Inject 75ul of 4μM target DNA [target DNA mimic the natural Herpes viral DNA, which binds to immobilized Probe 1]
6	Inject Taq buffer to rinse out non-attached target DNA
7	Inject 75ul of 4μM Probe 2 DNA [Probe 2 will bind to the other end of target DNA]
8	Inject Taq buffer to rinse out excess Probe 2 DNA

Table 1: Experimental procedures for DNA detection

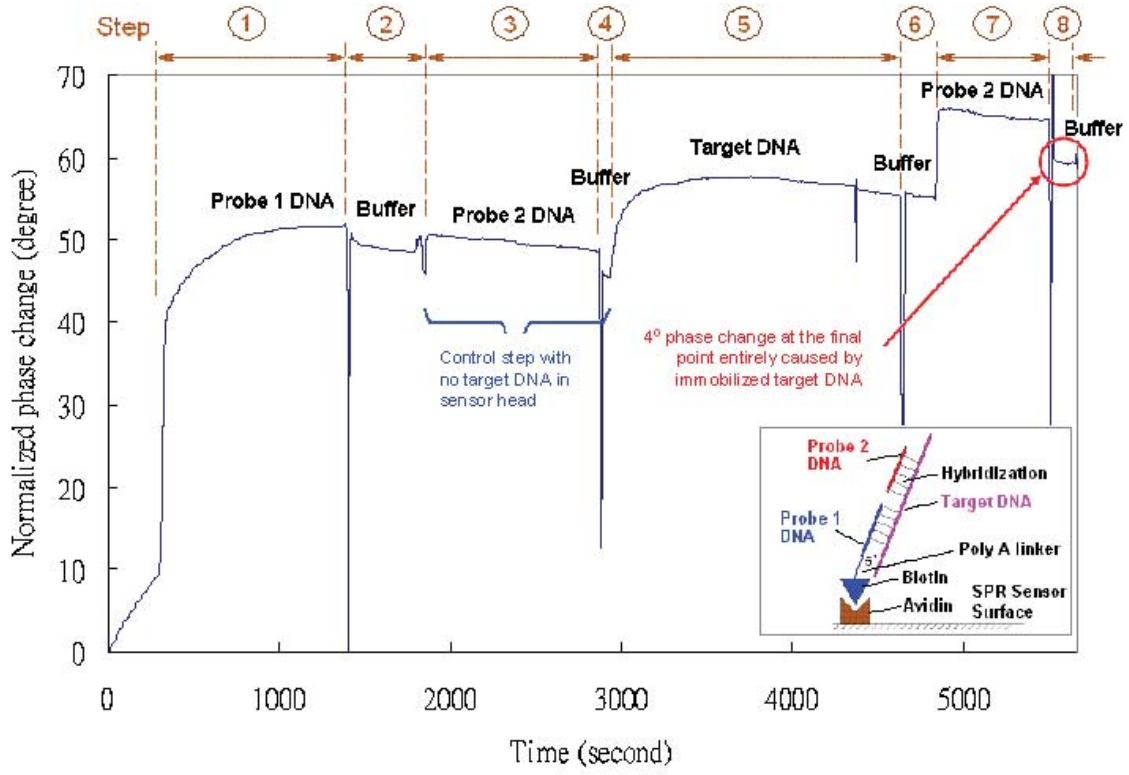


Figure 4: SPR phase plot showing the detection of Herpes DNA using complementary sequences Probe 1 and Probe 2. A net SPR phase change of 4 degrees caused by the immobilization of the target DNA has been observed.

As indicated in Figure 4, we have successfully demonstrated that our SPR-based system can perform label-free detection of Herpes DNA, thus opening the possibility for low-cost molecular diagnostics application of this technique. We are now in the process of combining the LCM technique with the image analysis system developed in the earlier part of this project for arrayed biosensing of practical biomolecular species in a clinical environment.

4. Publications and Awards

(May 2006 – Oct 2007)

[1] W. Yuan, H.P. Ho, Y.K. Suen, S.K. Kong, Chinlon Lin, "Improving sensitivity limit of surface plasmon resonance biosensors by detecting mixed interference signals", *Applied Optics* (accepted)

[2] C.L. Wong, H.P. Ho, Y.K. Suen, Winnie W.Y. Chow, S.Y. Wu, W.C. Law, W Yuan, W.J. Li, S.K. Kong and Chinlon Lin, "Two dimensional biosensor arrays based on SPR phase imaging", *Applied Optics*, 46(2007), 2325-2332.

[3] H.P. Ho, W. Yuan, C.L. Wong, S.Y. Wu, Y.K. Suen, S.K. Kong, Chinlon Lin, "Sensitivity enhancement based on application of multi-pass interferometry in phase-sensitive surface plasmon resonance biosensor", *Optics Communications*, 276(2007), 491-496.

[4] S.Y. Wu and H.P. Ho, "Single-beam self-referenced phase-sensitive surface plasmon resonance sensor with high detection resolution", *Chinese Optics Letters* (revised and under review)

[5] C.L. Wong, H.P. Ho, Y.K. Suen, S.Y. Wu, W. Yuan and S.K. Kong, "Two-dimensional phase surface plasmon resonance biosensor for protein array detection" (to be submitted to *Biosensors and Bioelectronics*)

[6] S.Y. Wu and H.P. Ho, "Highly sensitive single-beam phase-sensitive surface plasmon resonance biosensor with a wide dynamic range", *BIOS 2008*, in SPIE Photonics West, San Jose, USA, 19 - 24 January 2008.

[7] H.P. Ho, S.Y. Wu, C. Lin, S.K. Kong, "A simple differential phase imaging surface plasmon resonance sensor using liquid crystal modulator", *U.S. Patent* Pending.

[8] AWARD - 2nd runner-up of YDC Entrepreneurs' Challenge (E-Challenge) 2007 Business Plan Competition organized by The Young Entrepreneurs Development Council (YDC), 30 June 2007, Project title: "Surface Plasmon Resonance Biosensors".

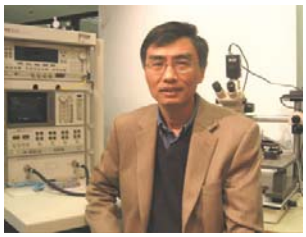
Bio-electromagnetic Modeling and Experiment Setup for Medical Electronics RF Safety Assessment

Principal Investigator:

Professor Ke-Li Wu [1]

Co-investigators:

Professor Pheng-Ann Heng [2]
Professor Ji Chen [3]



Research Team Members:

Dr. Lei Zhao, Research Associate [1]
Dr. Yougming Xie, Research Associate [1]
Mr. Xin Gao, Research Assistant [1]
Ms. Hongxia Li, Research Assistant [1]

[1] *Dept. of Electronic Engineering, CUHK*
[2] *Dept. of Computer Science and Engineering, CUHK*
[3] *Dept. of Electrical and Computer Engineering, University of Houston, USA*

Project Start Date: 1st June 2007

Abstract

In this project, we will develop the first 0.1 mm resolution electromagnetic human object model for accurate electromagnetic simulations, which will be the world finest virtual electromagnetic human model. A novel measurement and modeling approach to access the biomedical safety quantities during the operation of medical electronics will be established, and extensive simulation/modeling will be perform to understand biology effects due to external RF radiation at different operating frequencies. The proposed research work has significant impacts on both academic communities and industrial applications. Our interdisciplinary approach will open new areas of biomedical safety research and findings can be used in the future standard development in safety assessment. The success of this project will help to enhance the leading international position and to evaluate the reputation of the CUHK group in medical electronic safety assessment research and development. Most importantly, the research results will be used to guide the design of safe medical electronics.

1. Introduction

With rapid development in semiconductor technologies, miniature medical electronics are becoming an integral part of our daily life. Some of these devices, such as the medical telemetry systems, embed bio-sensors within human subjects and use radio-frequency (RF) technology to monitor patients on 24 hour basis, while other devices, such as the pacemaker or implanted cardioverter defibrillator, are being utilized to restore health and extend lives. In fact, the mobile phone is the most popular electronic device working in the proximity of our body. These devices

already have significantly impact the wellness of human beings and will have more impact on our lives in the future. As indicated by Luke Collins, “Medical electronics is set for rapid growth” in the near future [1].

Most medical devices are either implanted within human bodies or work in the proximity of human bodies. During the operation of these devices, electric signals will interact with surrounding human bodies as well as other near by electronic systems. Due to such interactions, these medical devices may also cause health concerns. For example, when medical telemetry systems transmit information back to hospital, just like typical mobile phone operations, some of the electromagnetic energies can be deposited into nearby human subjects. If the deposited energy exceeds a certain limit, it leads to many health concerns, such as thermal damage and brain cancer [2]-[4].

While radio-frequency radiation from medical devices raises safety concerns in human subjects, external electronic systems could cause more serious medical consequence for patients with implantable medical devices. For example, when patients with medical implants undergo a magnetic resonance imaging (MRI) examination, the MRI system can completely damage the functionality of implanted devices [5]. In addition, MRI system can cause significantly heating in tissues surrounding medical implants [6][7]. Similar concerns were raised when patients with implantable devices pass through airport walk-through-metal-detectors (WTMD) [8]-[12].

To ensure patient and public safety, the international electrotechnical commission (IEC) has developed a standard to limit the maximum energy deposition within human subjects under different environments [13][14]. These standards set the limits for electromagnetic related quantifies, such electric field, magnetic field, induced electric current, and specific absorption rate (SAR), for the whole body, partial body, and local region within human subjects. However, these standards were developed based on very crude models and have not considered complexity of real human subjects. Carefully studies on how external electromagnetic stimulus internal nerve systems were not performed. In addition, for some applications, such as the SAR measurement near implantable devices, direct measurement of SAR values near the implantable devices are impossible. Due to these limitations, international-wise, there is no recommended procedure to evaluate these quantities near the medical implantable devices.

Recognizing these facts, the international mobile manufacture forum (MMF) has initiated several research projects in a total amount of 2 million US dollar to seek for solutions. Although no significant research progress has been reported, the following issues have been identified. These issues include: 1) advanced enabling techniques for accurate safety assessment should be developed, 2) detailed human subject models need to be developed, 3) more complete biology effects of electromagnetic radiation needs to be studied, and 4) neuron stimulus due to electromagnetic radiation shall be investigated.

The goal of this project is to utilize the high-resolution human subject model developed at Chinese University of Hong Kong and develop enabling techniques for accurate RF safety assessment. With this these enabling technologies, two strong biomedical elements of the proposed research include 1) study the biology effects of electromagnetic radiations and 2) investigate the neuron stimulus due to external electromagnetic emissions. The completion of these tasks can significantly enhance the CUHK’s reputation in the international biomedical safety communities.

2. Signficance

The market for medical implants is a global one. The world-wide orthopaedic surgery market is approximate 9 billions US $ per year with an annual growth rate of approximate 5%. The current market for cardiovascular implants alone is approximate 6.4 billion US $ and for active implants approximate 5.3 billion US $ [14]. Millions of people’s well-beings are affected by these devices. The proposed research work has significant impacts on both academic communities and industrial applications. Our interdisciplinary approach will open new areas of biomedical safety research and findings can be used in the future standard development in safety assessment. The success of this project will help to enhance the leading international position and to evaluate the reputation of the CUHK group in medical electronic safety assessment research and development. Most importantly, the research results will be used to guide the design of safe medical electronics.

3. Methodologies and Objectives

In this proposed research, we will take the advantage of the complementary expertise and research experience of the co-instigators in biomedical safety research. Prof. Ke-Li Wu will oversee the development of this project as well as be responsible for the measurement tasks related to this research. Prof. Ji Chen will be responsible for the electromagnetic modeling aspects of the project and collaborate with Prof. Wu on developing the enabling technologies for safety assessment. Prof. Pheng-Ann Heng will collaborate with Prof. Wu on developing the first ever 0.1 mm resolution electromagnetic human model and collaborate with Prof. Chen on enhanced computational electromagnetic algorithms. Prof. Chen has authored a high frequency biomedical safety standard in IEEE ICES C95.3, 2006.

This project will be carried out via the completion of the following tasks:

(1) The creation of the world finest virtual electromagnetic human model, for bio-electromagnetics related researches, based on the 0.1 mm resolution virtual human model developed at CUHK. To develop the virtual EM human model, the current pixel based image model needs to be segmented into different organ objects so that the electromagnetic properties (i.e. complex permittivity) at different RF/microwave frequencies can be assigned to different tissues. This task requires a significant amount efforts from the researchers with medical background;

(2) Extensive simulation/modeling of RF exposure to the virtual EM human model for better understanding of biology effects due to external RF radiation at different operating frequencies; and

(3) The exploration of a novel fast modeling and testing approaches for medical device safety assessment. Since neither measurement nor modeling work alone can perform accurate safety assessment. The new approach will be able to predict the internal electromagnetic field strength based on the measurement data external to the human body and an efficient electromagnetic simulation.

4. References

[1] Luke Collins, "Medical dedication," *Electronic Systems and Software*, pp. 26-30, Vol. 4, Issue 5, October 2006.

[2] Lai, H., Horita, A., Chou, C.K. and Guy, A.W., "Low-level microwave irradiation affects central cholinergic activity in the rat," J. *Neurochem*, Vol. 48, pp. 40-45, 1987.

[3] Lai, H., Carino, M.A., Horita, "A. and Guy, A.W., Acute low-level microwave exposure and central cholinergic activity: a dose-response study," *Bioelectromagnetics*, Vol.10, pp. 203-209, 1989.

[4] Neil Cherry, "Evidence that Electromagnetic Radiation is Genotoxic: The implications for the epidemiology of cancer and cardiac, neurological and reproductive effects," presentation to NZ Parliament, June 2000.

[5] Fiek M, Remp T, Reithmann C, and Steinbeck G. "Complete loss of ICD programmability after magnetic resonance imaging." *PACE 2004*; 27:1002-4.

[6] Martin ET, Coman JA, Shellock FG, et al. "Magnetic resonance imaging and cardiac pacemaker safety at 1.5-Tesla," *Journal of the American College of Cardiology*, Vol. 43, pp. 1315-24, 2004.

[7] D. Wu, S. Shmsi, J. Chen, R. Liu, and W. Kainz, "Evaluations of Specific Absorption Rate and Temperature Rise Within Pregnant Woman Models in Magnetic Resonance Imaging Birdcage Coils," *IEEE Transactions on Microwave Theory and Techniques*, December, 2006.

[8] C. Kolb and et al., "Do airport metal detectors interfere with implantable pacemakers or cardioverter-defibrillators?" *J Am Coll Cardiol*, Vol.41, pp. 2054-2059, 2003.

[9] D. Wu, R. Qiang, J. Chen, W. Kainz, S. Sedman, "Safety Evaluation for Walk through metal detector," *invited for publication for IEEE EMC Newsletters*, 2005, to appear 2007.

[10] D. Wu, J. Chen, and W. Kainz, "Overexposure of pregnant women in walk through metal detectors according to ICNIRP Guidelines," *in proceedings of BEMS 2006*, June, Mexico, 2006.

[11] Dagang Wu, Ji Chen, David Jackson, Wolfgang Kainz, Howard Bassen, "SAR Evaluation for Medical Implants," *(invited) in proceeding of URSI meeting*, Boulder, CO, 2006.

[12] A. Christ, W. Kainz, J. Chen, and N. Kuster, "Current and future needs for the simulation of small and implanted antennas for medical applications," *in proceeding of IEEE international workshop on antenna technology, small antennas, and novel metamaterials*, March, 2006.

[13] International Electrotechnical Commission International standard, "medical equipment–part 2: particular requirements for the safety of magnetic resonance equipment for medical diagnosis," *International Electrotechnical Commission*, 60601-2-33, 2004.

[14] Institute of Electrical and Electronics Engineers (IEEE) C95.6TM 2002 IEEE Standard for Safety Levels with respect to Human Exposure to Radiofrequency Electromagnetic Fields, 0–3 kHz (New York: IEEE)

[15] http: isotc.iso.org, January, 2007.

Medical Applications of Terahertz Imaging

Principal Investigator:

Professor Emma MacPherson [1]

Co-investigators:

Professor Anil T. Ahuja [2]
Professor W.H. Cheung [3]
Professor K.T. Chan [1]
Professor Y.T. Zhang [1]

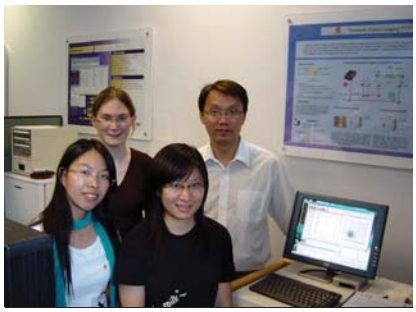
Collaborator:

Dr. Vincent Wallace [4]

Research Team Members:

Ms. Kanis Kan, Research Assistant [1]
Ms. Wing Sze Lee [3]

[1] *Dept. of Electronic Engineering, CUHK*
[2] *Dept. Diagnostic Radiology and Organ Imaging, CUHK*
[3] *Dept. Orthopaedics and Traumatology, CUHK*
[4] *TeraView Ltd, Cambridge, UK*



Project Start Date: 1st June 2007

Abstract

Terahertz (10^{12} Hz) pulsed imaging is a new technique and has only emerged in the last five years as a potential new clinical tool for medical imaging. The optical excitation required for emission of terahertz radiation is derived from a femtosecond pulsed laser (commonly a Ti-Sapphire laser at 800nm). The radiation produced is focused onto the sample of interest and then detected coherently in a reflection geometry such that the measurement is non-invasive. There are strong water absorptions in the terahertz region of the electromagnetic spectrum which therefore means that imaging using terahertz radiation would be a useful tool to investigate soft tissues. In this project we plan to investigate the terahertz properties of biological tissues with a view to determining the feasibility of using terahertz imaging in medicine. Our first case study under this project has been of rabbit knee cartilage – we are investigating the use of terahertz imaging to measure the variation in thickness of knee cartilage due to damage caused by osteoarthritis.

1. Introduction

Terahertz pulsed imaging (TPI™) is a non-destructive, non-ionising imaging modality with uses in medicine. TPI uses pulses of electromagnetic radiation typically with a full-width half-maximum of 0.3 ps and an average power of 100 nW. The pulses are detected coherently using a photoconductive device and the Fourier-transformed pulse gives a usable frequency range of 0.1–3 THz. A point measurement is analogous to an ultrasound A scan. Reflections off different layers are used to determine the structure at depth (similar to ultra-sound 'echoes').

By raster scanning to take several points within an area (C scan), TPI gives three-dimensional information and can reveal structures beneath the surface (with spatial resolution precision of 20um and axial resolution of 40um). A key advantage of terahertz imaging over techniques such as ultrasound and x-ray is that unique spectral information is obtained which can be used to distinguish between tissue types. Previous studies have revealed contrast in terahertz images of skin cancer and breast cancer [1]. Furthermore, terahertz spectroscopy data has shown that there are differences in the terahertz properties of these tissues which could account for the image contrast and potentially be used to determine the excision margins for tumours [2].

There are other areas in medicine in which terahertz imaging could be beneficial and have hitherto not been fully investigated. Osteoarthritis (OA) is one such example: OA is the most common form of arthritis, caused by the breakdown of cartilage. It usually affects weight-bearing joints like hip, knee, feet and spine, which causes the joints to degenerate and painful. After cartilage erosion, bone grinding may occur, leading to thickening and forming of osteophytes. As a result, pain, stiffness, swelling and reduced range of motion will be presented as symptoms. According to a survey in 2003, the estimated number of people with OA in Hong Kong was 200,000. Therefore, OA is a major medical challenge with high socioeconomic impact.

Early diagnosis of OA and subsequent prevention for retardation of OA and damage on subchondral bone is critical. Early diagnosis can help to treat joint problems effectively to prevent the onset of OA. Also, it will help to relieve pain, and improve mobility and quality of life. Clinically, the use of radiography for early OA diagnosis is very difficult. The classification of OA is also based on crude radiographic and clinical evidence only, as outlined by the American College of Rheumatology. Therefore, a promising quantitative approach to monitor articular cartilage is in pressing need. Recent work has shown that terahertz radiation can distinguish between normal and abnormal articular human cartilage highlighting the potential to use of terahertz imaging as a diagnostic tool for cartilage [3]. The Orthopaedics and Traumatology lab at CUHK has a well established osteoarthritis rabbit model - this is our chosen media to investigate the clinical application of terahertz imaging in this area.

2. Objectives and Significance

2.1. Vision

To develop terahertz imaging into a standard clinical tool for medical imaging and establish CUHK as an internationally recognised research institute for terahertz imaging

2.2. Key objectives

- To introduce terahertz imaging as a clinical tool in Hong Kong
- To conduct research on medical applications of terahertz imaging

2.3. Significance

Since terahertz imaging is a new development, many publications can be made simply on acquiring the terahertz properties of materials. At present, our main drive for research in terahertz is to investigate medical applications and in particular investigate the feasibility of terahertz probing in vivo and, through Prof MacPherson we are collaborating with TeraView Ltd to work with them to investigate medical applications such as determination of tumour excision margins, with a view to developing the world's first terahertz probe for in vivo use by clinicians. The faculties involved would then receive international recognition for this achievement.

3. Research Methodology

New Zealand white female rabbits, aged 27-32 weeks are used for OA model establishment according to our established protocol [4,5]. The rabbits are euthanized by overdose intravenous

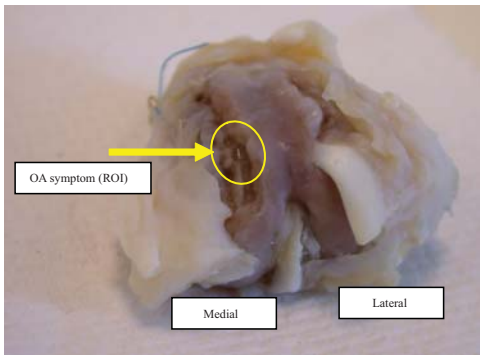


Figure 1: Femoral condyle with OA symptom

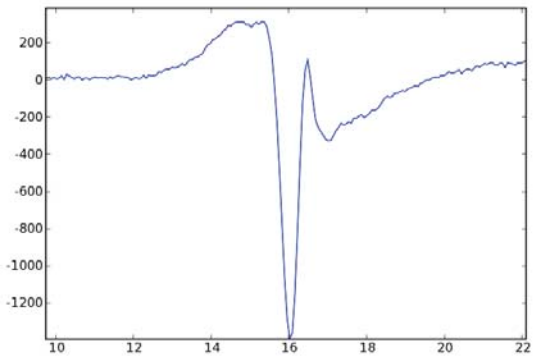


Figure 2: Typical reflected waveform from cartilage

barbiturate and both femoral knee joints are harvested and formalin fixed. One of the samples is illustrated below in Figure 1.

The areas of interest were placed on the terahertz imaging system (TPI Imaga 1000, TeraView Ltd). Gentle pressure was applied to make sure that the samples were making contact with the quartz imaging plate.

A typical reflected waveform from a region containing cartilage is illustrated in Figure 2. The first minimum (A) is the reflection of the terahertz light at the quartz/cartilage interface. The second minimum (B) is the reflection at the cartilage/bone interface. The optical delay between the minima is proportional to the thickness of cartilage. In Figure 3, we plot this optical delay as a colour map over the region of interest for one of the OA samples. The redder areas correspond to thicker cartilage and the lighter blue areas to thinner cartilage.

We plan to measure ten samples containing OA and ten control samples with no OA. We will then validate the terahertz depth measurement by using histomorphometry. Initially we will do this on the flat bed imaging system. We will then repeat the measurements when the THz probe is installed.

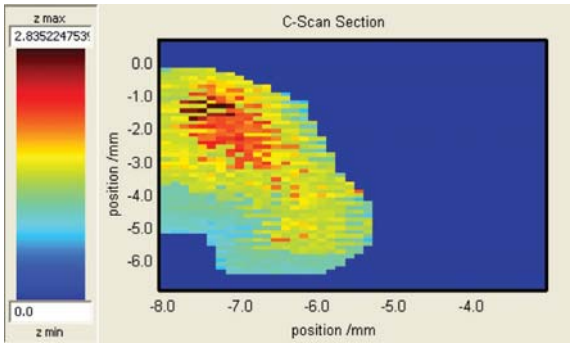


Figure 3: Colour map of knee cartilage thickness. The z-scale corresponds to the optical delay (in picoseconds) taken to pass through the cartilage layer.

4. References

[1] Fitzgerald A J, Wallace V P, Jimenez-Linan M, Bobrow L, Pye R J, Purushotham A D and Arnone D D, "Terahertz Pulsed Imaging of Human Breast Tumors", *2006, Radiology* 239 533–40

[2] Wallace V P, Fitzgerald A J, Pickwell E, Pye R J, Taday P F, Flanagan N and Ha T, "Ex vivo study of basal cell carcinoma using pulsed terahertz radiation", *J. Appl. Spectrosc.* 60(10), pp. 1127-1133(7)

[3] E. Jung, H. Park, J. Kim, Y. Han, H. Han, S. Kim, I. Park, J. Cui, B. Min, and H. Lim, "THz pulse imaging of human articular cartilage", *IRMMW-THz 2006 conference digest*, Shanghai, China.

[4] Fu LL, Maffulli N, Yip KM, Chan KM. "Articular cartilage lesions of the knee following immobilisation or destabilisation for 6 or 12 weeks in rabbits". *Clin Rheumatol.* 17(3):227-33, 1998.

[5] Fu LL, Maffulli N, Chan KM. "Intra-articular hyaluronic acid following knee immobilisation for 6 weeks in rabbits". *Clin Rheumatol.* 20(2):98-103, 2001.

Hybrid Assistive Knee Braces with Smart Actuators

Principal Investigator:

Professor Wei-Hsin Liao [1]

Co-investigators:

Professor Yunhui Liu [1]
Professor Ling Qin [2]
Professor Kai-Ming Chan [2]

Research Team Members:

Dr. Daihua Wang, Postdoctoral Fellow [1]
Dr. Zhehua Wu, Research Associate [1]
Mr. Jinzhou Chen, PhD student [1]
Mr. Hongtao Guo, PhD student [1]

[1] *Dept. of Mechanical and Automation Engineering, CUHK*
[2] *Dept. of Orthopaedics and Traumatology, CUHK*

Project Start Date: 1st June 2007

Abstract

Exoskeleton systems that can assist disabled people have been investigated in recent years. However, most exoskeletons use DC motors as actuators and batteries as power sources. While a DC motor consumes a lot of power and limits the working time of the exoskeleton, it produces large impedance that would cause discomfort and danger to people. Magnetorheological (MR) fluid is a smart fluid that can change its rheological behavior under applied magnetic field. Rotary devices using MR fluids have the benefits of high torque, good controllability, low power requirement, and safety. In this research, we develop a new MR actuator for leg exoskeleton. The MR actuator consists of a DC motor and an MR brake/clutch. When active torque is desired, the DC motor works and the MR actuator functions as a clutch to transfer the torque generated by the motor to the leg; when passive torque is desired, the DC motor is turned off and the MR actuator functions as a brake to provide controllable passive torque. The feasibility studies show that the adaptive control can track the needed torque/power of the knee joint very well, even under the variation of parameters.

1. Introduction

The percentage of aged persons in society is increasing, and their physical deterioration has become a socioeconomic problem in many countries, such as China, Japan, etc. In late 2004, the population of aged persons (older than 60 years) in China has reached 143 million. In this new century, the aged tendency of population will become more serious. It is estimated that the percentage of aged population would reach 40% in China between 2030 and 2050. Elderly people with weak muscle strength may not be able to walk frequently as before and would also lose their stability during walking. In turn, their muscles would be further deteriorated and they may become bedridden. The most effective means to avoid people from becoming bedridden is to provide a way for them to be able to maintain their physical activities.

Osteoarthritis (OA) is very prevalent in old people; it is about 60% in men and 70% in women after the age of 65 years [1]. Studies showed that adequate exercises would generate positive effects to OA patients. Some kinds of devices were developed to provide exercises; however, these devices are usually for hospital or home use and can't automatically provide assistance for people with mobility problems. Considering the large population base and the high percentage of elderly people, devices that can help people with weak muscle strength or OA will be in great need. Regarding this problem, researches were conducted to develop exoskeletons that can provide assistance for elderly and disabled people, such as RoboKnee developed by Pratt et al. [2], and Hybrid Assistive Limb (HAL) developed by Sankai et al. [3-4]. Although great progress has been made, these leg exoskeletons still have not been applied to people with mobility problems. Two major limiting factors remain for these exoskeletons to be used:

1) Safety: Current leg exoskeletons use sensors to detect human's intent and use motors to drive the exoskeletons to follow the intent of human body. Directly driving a motor to move a leg has the potential of danger, especial to people with weak muscles. Once the leg exoskeleton misunderstands the wearer's intent or goes out of order, the exoskeleton will cause the wearer to fall.

2) Short life: After a full charge, the RoboKnee can only work 30-60 minutes in heavy use. As the most successful exoskeleton up to now, a fully charged HAL system lasts for 2 hours and 40 minutes. Therefore, the life time is insufficient for practical applications.

Magnetorheological (MR) fluids comprise microscale ferromagnetic particulates dispersed in a carrier liquid. When a magnetic field is applied to this fluid, the dispersed magnetic particles will line up along the direction of the magnetic field. A shear stress is needed to disrupt this chain structure. The apparent yield stress increases as the magnetic field increases. As a smart fluid, MR fluid has several advantages such as high yield stress, good stability, and fast response time. These features make the MR fluid promising for various applications. With the use of MR fluid, an MR damper was used in a rehabilitation device to provide controllable resistance with excellent force tracking performance [5].

2. Objectives and Significance

In this project, a new leg exoskeleton will be developed utilizing an MR actuator. The idea is to use the MR actuator to provide controllable assistive torque to the knee joint and thus make the knee brace adaptable to the person while interacting with environment. The MR device can function as a clutch or a brake while the motor will be turned on when needed. Comparing with previous leg exoskeletons, the new hybrid assistive knee brace with MR actuator will have the following benefits:

1) Better safety: While functioning as clutch, the MR device transfers the torque generated by the motor to the human body. Using MR device as clutch, the inertia of output shaft can be made very small and low impedance can be achieved, thus the system is much safer than the existing exoskeleton, which assistive torque is directly provided by the motor.
2) Longer life: While functioning as brake, the MR device can produce large torque while consuming little power. A prosthetic leg using an MR damper powered by a small Li-ion battery can work about 2 days between two charges [6]. For comparison, the RoboKnee can only work 30-60 minutes even the battery used is much larger. Obviously, MR devices are more energy efficient.

The new hybrid assistive knee brace with MR actuator been developed will have the following advantages: compact in size, low power consumption, safe, and adaptive. It will provide the disabled people with user-adaptive knee braces while increasing the level of biological realism compared to passive knee braces.

3. Research Methodology

The configuration of a leg exoskeleton with an MR actuator is shown in Fig. 1, in which the controller and power supply are not shown. The main parts of the system include: braces, MR fluid actuator, sensors. The braces include upper brace and lower brace. The upper brace is bound to the upper leg and the lower brace is bound to the lower leg. The lower brace is connected to the foot, and the upper brace is connected to the waist (the upper part of the upper brace is not shown in Fig. 1). The braces transfer the assistive torque generated by the actuator to the lower limb. An elastic element is attached to the bottom of the shoe to absorb shock, thus provides protection to the system. The sensors are used to detect the user's walking condition and estimate the needed assistive torque. Two force sensors are mounted on the front and rear sole of the foot to measure the reaction force from the floor. Two strain gages are mounted on the aft and fore of the lower brace to measure the force acting on the actuator. The torque produced by the actuator can be measured using these strain gages. An angular sensor is to measure the knee joint angle. The knee angular velocity can also be derived by differentiation of the angle signal.

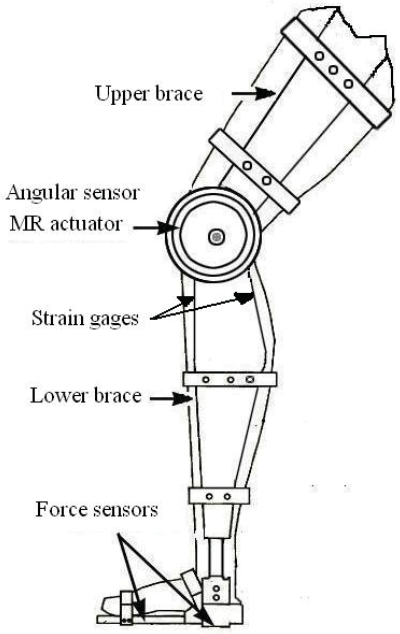


Fig. 1 Configuration of a leg exoskeleton with an MR actuator

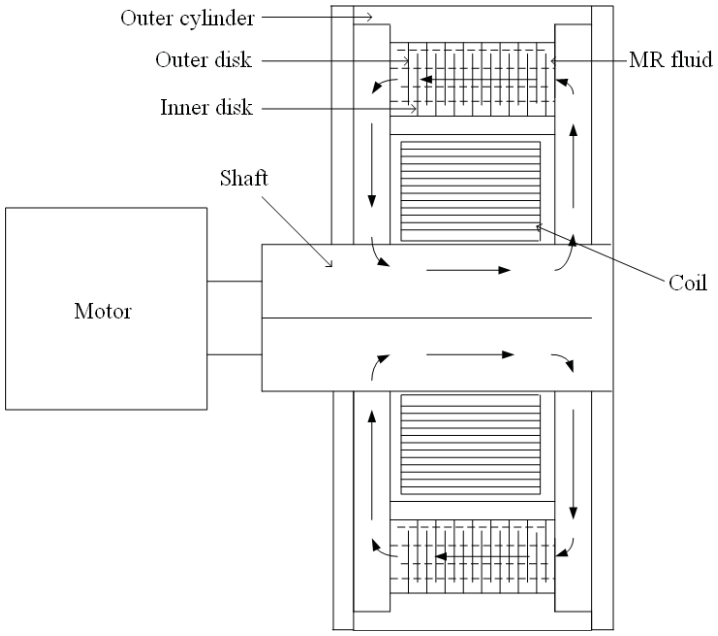


Fig. 2 Schematic diagram of actuator with MR fluid

The structure of the MR fluid actuator is shown in Fig. 2. MR fluids are used in shear mode to produce or transfer torque. The outer cylinder is connected to the upper leg, the shaft is connected to the motor, and the motor is mounted on the lower brace. The shaft can be locked to the lower brace. For this actuator, there are three working conditions:

1) The shaft is locked to the lower brace and the magnetic field is on. The MR device acts as a brake, which can provide controllable passive torque.

2) The magnetic field is off. The MR device doesn't work and the knee joint can rotate freely.

3) The shaft is unlocked to the lower brace and the magnetic field is on, while the motor is working. As the shaft is unlocked to the lower brace, the shaft can rotate relatively to the lower leg, thus the MR device acts as a clutch to transfer the torque from the motor to the upper leg.

The main functional components are the alternately placed outer disks (or stators) and inner disks (or rotors). The outer disks are mounted to the outer cylinder by splines. The inner disks are connected to the shaft by inner splines, which can rotate relatively to the outer disks as the shaft rotates. MR fluids are filled in the gaps between outer disks and the inner disks. As the shaft rotates and the magnetic field is on, the torque due to shear effect will be generated between the outer and inner disks. This torque is controlled by the applied magnetic field.

4. References

[1] P. Sarzi-Puttini, M.A. Cimmino, R. Scarpa, R. Caporali, F. Parazzini, A. Zaninelli, F. Atzeni, and B. Canesi, "Osteoarthritis: An Overview of the Disease and Its Treatment Strategies," *Seminars in Arthritis and Rheummatism*, 35 (1), pp. 1-10, 2005.

[2] J.E. Pratt, B.T. Krupp, C.J. Morse, and S.H. Collins, "The RoboKnee: An Exoskeleton for Enhancing Strength and Endurance During Walking," *Proceedings of the IEEE International Conference on Robotics and Automation*, pp. 2430-2435, 2004.

[3] H. Kawamoto and Y. Sankai, "Comfortable Power Assist Control Method for Walking Aid by HAL-3," *IEEE International Conference on Systems, Man and Cybernetics*, vol. 4, pp. 190-193, 2002.

[4] E. Guizzo and H. Goldstein, "The Rise of the Body Bots," *IEEE Spectrum*, pp. 50-56, October 2005.

[5] S.F. Dong, K.Q. Lu, J.Q. Sun, and K. Rudolph, "Rehabilitation Device with Variable Resistance and Intelligent Control," *Medical Engineering and Physics*, vol. 27, pp. 249-255, 2005.

[6] J.D. Carlson, W. Matthis, and J.R. Toscano, "Smart Prosthetics based on Magnetorheological Fluids," *Proceedings of SPIE Conference on Smart Structures and Materials*, vol. 4332, pp. 308-316, 2001.

Research on Language and Brain Waves

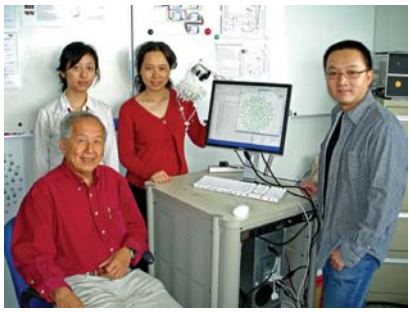
Principal Investigator:

Professor William S.-Y. Wang [(1)]

Research Team Members:

Dr. James W. Minett, Postdoctoral Fellow [(1)]
Ms. Lan Shuai, PhD student [(1)]
Ms. Hong-Ying Zheng, PhD student [(1)]
Mr. Francis C.-K. Wong, PhD student [(1)]

[(1)] *Dept. of Electronic Engineering, CUHK*



Project Start Date: 1st September 2007

Abstract

The central goal of this project is to understand the biological and cultural bases of language in terms of the underlying mental activities. More specifically, our goal is to study linguistic behaviour with particular reference to the distinctive features of the Chinese language.

1. Introduction

We look into brain activities that underlie linguistic behaviour through the electroencephalograph (EEG). The time-locked brain waves become event related potentials (ERPs), through which we study the human brain in action in a non-invasive way [1].

The several aspects of linguistic behaviour currently under investigation, partially supported by SHIAE, include:

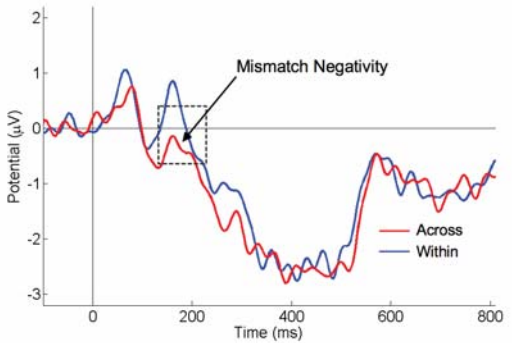
- Perception of tonal syllables, as well as the lateralisation of the brain activities involved.
- Similarities and differences, in terms of brain activities, in perceiving speech versus non-speech sounds.
- Perception of meaning in sentential contexts.

2. Experiment on Categorical Perception of Speech

The categorical nature of perception in the context of Chinese tonal syllables has long been our interest [2]. Though the tonal contours of a syllable can vary physically continuously between a low frequency tone (e.g. '罵' in Cantonese) and a high frequency one (e.g. '媽' in Cantonese), native speakers often perceive the continuum as a series of discrete categories. As a consequence, when asked to discriminate pairs of stimuli, the performance for pairs that cross a categorical boundary is much better than that for within category pairs.

As an extension to our behavioural studies reported in [3, 4], we are interested in zooming in to look at the time course to address the issue of when the perception of linguistic categories arises. A pilot ERP study has been conducted by Zheng Hong-Ying using an odd-ball passive listening paradigm in which she looks at the brain responses to auditory stimuli that deviate from standard ones. The ERP component named Mismatch Negativity (MMN), which has become standard in

neurocognitive studies [5], is examined. Preliminary results, plotted on the right, show that MMN for across-category deviants (red line) is more negative than within-category ones (blue line) even when the subjects' focal attention is drawn elsewhere. We are currently carrying out statistical analysis of the data to fine tune the experimental design for a full-scale study. We will also include subjects who are not speakers of Chinese to study the influence of language experience on the brain.



ERP, grand-average for eight subjects, obtained from categorical perception of speech.

3. Experiment on Lateralisation of Brain Activities

The lateralisation of brain activities for tone perception has been debated for many years [6-9]. A recent study that elucidated the time course of the brain waves involved was reported in *PNAS* [5]. Other than the MMN component, which peaks at around 200ms, we would also like to explore the early and late components such as N1 (which peaks at around 100ms) and N400 (which peaks at around 400ms). If the perception of tones undergoes an integration process from physical properties to meaning contrasts, then it will be expected that the lateralisation of the corresponding brain activities to occur in the right hemisphere during the early stages of processing and in the left hemisphere during the later stages. Shuai Lan is conducting experiments, the preliminary results of which show that the lateralisation of the N1 component for tones in dichotic listening tasks is significantly different from that for consonants. Her behavioural experiments [10] show also that the lateralisation pattern for lexical tones is mediated by the signal-to-noise ratio of the stimuli as well as the physical contours of the lexical tones.

4. Experiment on Comprehension

A third study that we are planning concerns the comprehension of Chinese. The motivation comes from a study published in *Science* in which the brain responses to deviant words in sentence contexts were investigated [11]. For instance, the N400 ERP component elicited by deviant words (e.g., 'gravity' and 'China' in S2 and S3, respectively, below) in reading visually presented sentences was compared with that for words that do not violate semantically the sentence context (e.g., 'Germany' in S1).

S1:	*'The fall of the Berlin Wall reunited **Germany**'*	Congruent sentence
S2:	*'The fall of the Berlin Wall reunited **gravity**'*	Word-meaning violation
S3:	*'The fall of the Berlin Wall reunited **China**'*	World-knowledge violation

It was found that both word meaning violation and world knowledge violation elicit an N400 greater in magnitude than that elicited by congruent sentences. Further analysis using time-frequency techniques and wavelet analysis [11, 12] also revealed possible differences in the spatial sources of neural generators among the three conditions.

We are interested in carrying out a parallel study in the context of Chinese reading in light of the distinctive features of Chinese, such as its use of a morpho-syllabic writing system, the abundance of homophones, and the large degree of ambiguity. Examples of materials in Cantonese that we have prepared for the pilot study, conducted in early November, are shown below.

C1:	吸煙	會	導致	肺癌	同埋	心臟病	Congruent sentence
	smoking	will	cause	lung cancer	and	heart disease	
C2:	吸煙	會	導致	肺癌	同埋	汽水罐	Word-meaning violation
	smoking	will	cause	lung cancer	and	soft-drink can	
C3:	吸煙	會	導致	肺癌	同埋	糖尿病	World-knowledge violation
	smoking	will	cause	lung cancer	and	diabetes	

5. Research Methodology

Our main research tool is a 128-channel Geodesic EEG System (GES 250) manufactured by Electrical Geodesics, Inc. The equipment was moved from the Prince of Wales Hospital to Ho Sin-Hang Engineering Building in the Summer of 2007. Given the newness of our project, we have decided to pursue a dual path approach in our work: theory and experiment.

In the theory part, we are reviewing the relevant literature on the neuroscience of brain waves. This literature includes various approaches to understanding the generators of these waves, methods for their analysis, and reports of experiments done elsewhere. Concurrently, we are exploring the use of wavelet analysis, chaos theory, as well as becoming familiar with the software packages we have acquired for source localisation, e.g., sLoretta and BESA.

In the experiment part, we are building upon our previous behavioural studies to look for neural correlates associated with observed linguistic behaviours. We search for potential breakthroughs in the study of the biological bases of language. Our first step towards this end is to conduct studies parallel to well established ones, such as [5, 11, 12] mentioned above, in the context of the Chinese language.

6. Activities Undertaken

Setup of the EEG system

The group from which we have acquired the GES 250 utilised it mainly for clinical purposes. After the transfer of ownership to our side and the relocation to our space, upgrade and maintenance have been carried out. They include purchase and installation of:

- Mac and Windows based computers, for data acquisition and experimental control, respectively.
- Essential software packages, such as Net Station, E-Prime, BESA and sLoretta.
- Single-clock hardware and software package, for achieving precise timing synchronisation.

As technical support for GES 250 is virtually non-existent in Hong Kong, our team has been making an effort to understand the processes involved in setting up the system. Currently, our EEG system is in a functional condition, and we have started making use of it to gather data.

Overseas training

In the late summer of 2007, Francis Wong was sent to the University of South Florida, USA. He was given a 2-week training provided by the research team of Prof. Emanuel Donchin, who has been one of the leaders in EEG research for the past several decades. More importantly, the laboratory of Prof. Donchin is equipped with the same system as ours. During the trip to Florida, Francis Wong had the chance to seek advice on conducting EEG experiments and setting up equipment and protocols for our laboratory, and was given a tutorial on basic statistical analysis of EEG data.

In-house tutorials

Dr. James Minett discussed with the team the use of time-frequency techniques for analysis of EEG waves. The discussion, held on 17th October, covered the bases of Fourier analysis, as well as the comparison of it with Hilbert and wavelet analysis. On 1st November, he led another discussion on looking at brain activities from the perspective of chaos theory, an approach first advocated by Walter J. Freeman [13].

On 7th November, Prof. Ron Chen, director of the Centre for Chaos and Complex Networks at City University of Hong Kong gave a tutorial to the team on the study of brain waves as chaotic signals.

Exchanges with collaborators

We have been interacting on a regular basis with Dr. Yang Chin Lung and Dr. Peng Gang from the State Key Laboratory of Brain and Cognitive Sciences, The University of Hong Kong. Dr. Yang presented his EEG/ERP work on Chinese sentence processing to us to establish a basis for future collaboration. Dr. Peng is an expert in signal processing, and he has been advising us on the possible uses of time-frequency analysis of EEG data. On 6th December 2007, the PI will co-organize, in collaboration with Prof. Tan Li-hai, director of the State Key Laboratory of Brain and Cognitive Sciences, HKU, a public lecture at HKU by Prof. Paul Kay of the University of California, Berkeley.

Workshops, conferences and invited lectures

On 2nd October 2007, the PI lectured at the Brain Sciences Institute of RIKEN, Tokyo, and discussed possible collaboration with Prof. Atsushi Iriki and Prof. Kazuo Okanoya, both of RIKEN.

On 4th October, the PI met with Prof. Junjiro Kanamori, director of the International Institute of Advanced Studies in Kyoto, to plan an international seminar there in April 2008. The theme of the workshop is *Language and The Brain*. It is expected that our team will learn the latest results from leading researchers as well as present our own research results there.

The PI has been invited to the following international conferences for presentations during the Summer of 2008:

- BRAIN TALK: DISCOURSE WITH AND IN THE BRAIN. 2-3 June 2008 at Lund University, Sweden. Organized by Lund University, Newcastle University, and University of Oslo.
- BIOLOGICAL FOUNDATIONS AND THE ORIGIN OF SYNTAX. 13-18 July 2008 at Ernst Strungmann Forum in Frankfurt, Germany. Chaired by Derek Bickerton and Eörs Szathmáry.

7. References

[1] E. Donchin, "fMRI: Not the only way to look at the human brain," *Observer*, vol. 19, 2006.

[2] W. S.-Y. Wang, "Language change," *Annals of the New York Academy of Science*, vol. 280, pp. 61-72, 1976.

[3] H. Y. Zheng, G. Peng, P. W.-M. Tsang, and W. S.-Y. Wang, "Perception of Cantonese level tones influenced by context position," in *3rd International Conference on Speech prosody*, Dresden, Germany, 2006.

[4] H. Y. Zheng, P. W.-M. Tsang, and W. S.-Y. Wang, "Categorical perception of Cantonese tones in context: A cross-linguistic study," in *10th European Conference on Speech Communication and Technology (Interspeech 2007)*, Antwerp, Belgium, 2007.

[5] H. Luo, J.-T. Ni, Z.-H. Li, X.-O. Li, D.-R. Zhang, F.-G. Zeng, and L. Chen, "Opposite patterns of hemisphere dominance for early auditory processing of lexical tones and consonants," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, pp. 19558-19563, 2006.

[6] H. L. Jamison, K. E. Watkins, D. V. M. Bishop, and P. M. Matthews, "Hemispheric specialization for processing auditory nonspeech stimuli," *Cerebral Cortex*, vol. 16, pp. 1266-1275, 2006.

[7] J. J. Sidtis, "The complex tone test: implications for the assessment of auditory laterality effects," *Neuropsychologia*, vol. 19, pp. 103-112, 1981.

[8] D. Van Lancker and V. A. Fromkin, "Hemispheric specialization for pitch and 'tone': Evidence from Thai," *Journal of Phonetics*, vol. 1, pp. 101-109, 1973.

[9] Y. Wang, A. Jongman, and J. A. Sereno, "Dichotic perception of Mandarin tones by Chinese and American listeners," *Brain and Language*, vol. 78, pp. 332-348, 2001.

[10] L. Shuai and W. S.-Y. Wang, "Tone lateralization under noisy conditions," in *The 4th International Conference on Speech Prosody* Campinas, Brazil, in-press.

[11] P. Hagoort, L. Hald, M. Bastiaansen, and K. M. Petersson, "Integration of word meaning and world knowledge in language comprehension," *Science*, vol. 304, pp. 438-441, 2004.

[12] L. A. Hald, M. C. M. Bastiaansen, and P. Hagoort, "EEG theta and gamma responses to semantic violations in online sentence processing," *Brain and Language*, vol. 96, pp. 90-105, 2006.

[13] W. J. Freeman, *Neurodynamics: An exploration in mesoscopic brain dynamics*. London: Springer-Verlag, 2000.

Multimedia Technologies

We are living in an exciting era that benefit from the rapid changes being unleashed by the Internet and the information technology. Armed with more computational power and faster communication networks, scientists and engineers can now create more exciting applications and services to end users. Some of the applications or services that we thought only exist in the science-fiction movies are now within grasp and may become reality within this decade.

Multimedia technologies constitute a multi-disciplinary field that extends beyond engineering to reach the arts, sciences and humanities. Main contributive areas include engineering, computer science, artificial intelligence, communications, design, ergonomics and human factors, cognitive psychology, linguistics, and philosophy. The impact of multimedia applications is tremendous – it can range from medicine to entertainment; from bridging the societal digital divide to remote space exploration. In short, our imagination is the limit.

At the Faculty of Engineering, many professors and students are dedicated to advancing the frontiers of multimedia technology. One can easily find a plethora of relevant research projects and home grown technologies that can benefit different types of end users and companies. The research teams have acquired more than $58 million from the Industry Support Fund and Innovation and Technology Fund over the past decade to support various kinds of multimedia research and development projects, a strong indication of the strength, as well as our commitment in this important area.

“What does the future hold with these advanced multimedia technologies?” A question usually put forth by end users. Rather than making a futuristic projection, the teams prefer the practical approach taken by Dr. Alan Kay, a Turing award winner, who said: *“the best way to predict the future is to invent one”*. We strongly believe that it is an exciting journey that will vastly improve the way to look and process information in the coming decades, and we hope that interested parties can join us in this endeavor. Currently, we are working on interesting topics including:

1. Audiovisuals: photorealistic graphical animation, multilingual speech recognition and synthesis, avatars.

2. Communications: multimedia networking, compression, wireless communication.

3. Digital Libraries: content-based image, video and audio retrieval, information visualization.

4. Games: three dimensional game engines, mobile and multi-player games with parallel rendering methods.

5. Human-computer interaction: human-centric, mobile and multimodal interfaces, human-robot interaction, wearable computing, virtual reality, augmented reality, biometrics-secured computing

Research Reports in **Multimedia Technologies**

Completed Projects

- **Mobile Wireless Multimedia Communication**
- **An Automatic Multi-layer Video Content Classification Framework**
- **Automatic Multimedia Fission, Categorization and Fusion for Personalized Visualization in Multimedia Information Retrieval**

On-going Projects

- **Automatic Video Segmentation and Tracking for Real Time Multimedia Services**
- **Information Retrieval from Mixed-Language Spoken Documents**
- **Wireless Mesh Network Testbed**
- **Real-time Transmission of High Definition (HD) 3D Video and HD Audio in Gigabit-LAN**
- **High Dynamic Range Image Compression and Display**
- **Multimedia Content Distribution over Hybrid Satellite-Terrestrial Communication Networks**

Mobile Wireless Multimedia Communication

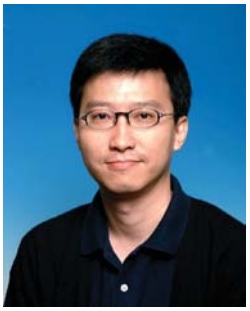
Principal Investigator:

Professor John C.S. Lui [1]

Co-investigator:

Professor Dah-Ming Chiu [2]

[1] *Dept. of Computer Science and Engineering, CUHK*
[2] *Dept. of Information Engineering, CUHK*



Project Start Date: 1st March 2005
Completion Date: 28th February 2007

Under the Mobile Wireless Multimedia Communication Research Project, there were two main initiatives, namely:

- Wireless Multimedia Communication Systems,
- Multi-player Game Applications.

In the following, I report the major achievements, such as scholastic outputs, as well as prototype systems (both hardware and software) that we have built.

Wireless Multimedia Communication Systems

One important aspect on the wireless communication systems is to understand the fundamental working principle of IEEE 802.11 multi-hop ad-hoc wireless networks. In wireless networks, it is well understood what throughput can be achieved by nodes who can hear each other (i.e. nodes within a single cell). The effects of nodes beyond the sensing range (known as hidden nodes) on a sender are complicated and difficult to analyze. Consequently, how to analytically model multi-hop ad-hoc networks, specially networks based on the popular IEEE802.11 standards remains largely open. In this research, we provide a unified model to characterize different types of hidden node interference and the analysis of 802.11 wireless networks, and derive a general solution for throughput given a linear network of *arbitrary density* and *transmission distance* between source and destination nodes. An important insight from our result is that there is a certain transmission distance, which is less than the maximum transmission distance that optimizes throughput in such networks.

In our second work, we present a methodology to analytically compute the *throughput capacity*, or the maximum end-to-end throughput of a given source and destination pair in a multi-hop wireless network. The end-to-end throughput capacity is computed by considering the interference due to neighboring nodes, as well as various modes of hidden node interference. Knowing the throughput capacity is important because it facilitates the design of routing policy, admission control for realtime traffic, as well as load control for wireless networks. We model location-dependent neighboring interference and we use a contention graph to represent these interference relationships. Based on the contention graph, we formulate the individual link capacity as a set of fixed point equations. The end-to-end throughput capacity can then be determined once these link capacities are obtained. To illustrate the utility of our proposed methodology, we present two important applications: (a) *route optimization* to determine the path with the maximum end-to-end throughput capacity and, (b) *optimal offered load control* for a given path so that the maximum end-to-end capacity can be achieved.

Some of the scholastic outputs are:

- "Determining the End-to-end Throughput Capacity in Multi-Hop Networks: Methodology and Applications". *ACM Sigmetrics/Performance Conference*, 2006. **(Acceptance rate: 30/217=14%)**

- "The fundamental role of hop distance in IEEE802.11 multi-hop Ad-Hoc wireless networks". International Conference of Network Protocols (ICNP), 2005. **(Acceptance rate: 36/212=17%).**

- We have extended the above work and submitted the paper for journal publication.

In summary, these three pieces of work have attracted many attention in the wireless network community since it provides a systematic approach to determine an end-to-end capacity, one can easily build many multimedia applications (e.g., video/audio streaming) on this type of network.

Multi-player Game Applications

Another major theme of our project is to come up with effective mean to securely distributed large video segments for different players, as well as a scalable and efficient method to detect cheats (which can occurred in many different forms) for massively multi-player games. For video streaming, we introduce a novel concept of cooperative proxy-caching system, and we show the scalability and high performance of the system. For the cheat detection, we provide an effective mean to detect aimbot cheats. The detection method is based on the dynamic Bayesian network approach. Also, our detection framework relies solely on the game states and runs in the game server only. Therefore it is invulnerable to hacks and it is a much more deployable solution. Some of the scholastic outputs are:

- "Detecting Cheaters for Multi-player Games: Theory, Design and Implementation". *Second IEEE International Workshop on Networking Issues in Multimedia Entertainment* (NIME'06). Las Vegas, USA, January, 2006. (An extended version has been accepted by the International Journal of Network Security).

- "The Design and Analysis of the Secure Multimedia Library".
  Accepted for publication in the ACM Multimedia Systems Journal.

- "COPACC: An Architecture of Cooperative Proxy-Client Caching System for On-Demand Media Streaming". *IEEE Transaction on Parallel and Distributed Systems, 18(1), 70-83, January, 2007*.

- "A Revenue-Rewarding Scheme For Cooperative Proxy Media Streaming Systems". Accepted for publication in the ACM Transactions on Multimedia Computing, Communications and Applications.

Note that we also built a software prototypes to (a) detect cheats, and (b) provide secured group communication and proxy services for multi-players games. For more information, please refer to the released software under

http://www.cse.cuhk.edu.hk/~cslui/ANSRlab/software.html

Due to availability of other external fundings, we decided to return the unspent money back to SHIAE in 2007 so that SHIAE can fund other worthwhile projects.

Currently, we are also in the process of negotiating licensing fee to several network game companies, both in China and Hong Kong. In the past month, OrbisIP/ICIPR from England is interested in licensing our technologies. We are going through the negotiation at time time.

An Automatic Multi-layer Video Content Classification Framework

Principal Investigator:

Professor Michael Rung-Tsong Lyu [1]

Research Team Members:

Dr. Steven Chu-Hong Hoi, Project leader and PhD graduate [1]
(Currently he is an Assistant Professor at National Technological University of Singapore)
Mr. Jianke Zhu, PhD candidate [1]
Mr. Wyman Wong, MPh graduate [1]
(Currently he works for Fujitsu, HK)

[1] *Dept. of Computer Science and Engineering, CUHK*

Project Start Date: 1st March 2005
Completion Date: 31st August 2007

Project Website:
http://www.cse.cuhk.edu.hk/~lyu/SHIAE

Abstract

Content-based classification of video stories enables efficient indexing of video information, allowing users to locate their desired stories faster and more accurately. Emerging techniques for extracting, representing and understanding video structures for content browsing and retrieval are of acute interest and timely importance in managing vast video contents. However, most video classification methods still use texts, including captions and speech scripts, as the only clue to index video contents. Since keyword-based techniques cannot adequately represent the semantic information in video, this project proposes an innovative approach to classify a video story by both syntactic features and semantic features under a generic framework. We first investigate the appropriate definition of these features with appropriate extraction schemes. In addition to the textual information, aural and visual information is also considered in generating the Video Content Feature (VCF). A number of classification models are analyzed to construct an appropriate VCF classifier. The classifiers are trained to generate a multi-layer class hierarchy, which is extendable to include new story classes. After the classification, the video story is organized into a tree structure to support a category-based query mechanism. We also build intra- and/or inter-class relationship among video stories, enabling rich representation and indexing of the video contents. Modern machine learning techniques are also investigated, including Nearest Neighbor (NN) Search, Bayesian classifiers, Support Vector Machines (SVM), Supervised Learning, Semi-Supervised Learning, Collaborative Learning, and Active Learning. Finally, we build a large-scale image/video information retrieval system to encapsulate our research schemes.

1. Introduction

1.1. Background of the project

Providing users with their desired video stories as fast as possible is the ultimate goal of a video library system. Usually, users come to a video library with an idea about the type of stories they want, but they may not be sure of what words will definitely and frequently appear in the stories. Due to the keyword sensitivity and the global searching scope, in a conventional keyword-based video library, two scenarios often happen after the user enters the keywords: (1) no results are returned, and (2) a number of results are returned, but most of them are undesired. To enable the users to get their desired stories quickly, we seek to provide users with a category-based video library so that they can enter the related category to find what they want. The prerequisite of building such a video library is to automatically classify video stories by their contents with modern machine learning techniques.

The automatic video content classification is currently an open problem. This project investigates the whole process beginning with a segmented video story. We attempt to answer the following questions:

(1) What is the content of a video story?
(2) How to extract the content of the story?
(3) How to represent the content of the story to make it distinctive?
(4) How to classify the type of content?

Generally, the answer to the first question is the perceptible aural, visual and textual information carried by the video story. But how much and to which level will this information be used is different from method to method. In this project, we utilize both syntactic and semantic features to address Question 1. The former includes many low-level features, while the latter includes domain-specific and knowledge-based features. To extract these features for Question 2, a number of images, video and audio processing algorithms are employed or developed. The key issues include video text recognition, speech recognition, audio classification, logo detection, face recognition, and video object detection and tracking. We address these issues in the research project. Question 3 is the most important issue that determines the accuracy of content classification. Since bridging the gap between syntactic and semantic features in the current state-of-the-art is difficult, this project investigates whether the syntactic features and semantic features can be integrated into a generic Video Content Feature (VCF) definition. Each story is then mapped to a high-dimension VCF vector. When the gap is too wide and such integration is deemed difficult, the syntactic features and the semantic features must be represented separately. The solution to Question 4 actually depends on the answer to Question 3. If a VCF vector is available, we can design a multi-layer classification network based on neural network or Bayesian classifier. After training with a number of sample stories, the VCF vectors of the same story type are clustered in the feature space, which becomes the characteristic VCF vector of this story type. Otherwise, the syntactic features have to be processed by the classification network and the semantic features by the knowledge-based reasoning network, and finally the information fusion techniques are brought in to produce the classification results.

According to the classification results, all stories are well organized into a tree structure to support the category-based query. To enrich the delivery format, the intra- and/or inter-category relationships based on timing, topic or geographical information are built to support hyper links, various query conditions, and story series on the same theme. Since the work in this project belongs to the backend processing of a video library, we can employ a proper distributed computing model to guarantee the work to be done smoothly and efficiently. This distributed system implementation is achieved by a scalable content-based image retrieval scheme using the locality-sensitive hashing (LSH) technique.

1.2. Objectives of the project

The major objectives in this project are therefore summarized as follows:

1. Syntactic feature and semantic feature definition criteria and extraction algorithms
2. Visual feature extraction from image sequences of video
3. Multi-layer video content class hierarchy mechanisms
4. Unified learning framework for large-scale multimedia classification problems
5. A variety of machine learning algorithms for video content classification and retrieval
6. Experimental system implementation for content based video information retrieval.

1.3. Significance of the project

A number of significant results are obtained in the project. First, the automatic multi-layer video content classification method contributes to video content analysis research. Although this method is not versatile for all kinds of story structures in different countries, it can be trainable and extendable to different story structures. Secondly, the category-based query is an innovation for video libraries. Since the searching scope is diminished to the specified category, users can enjoy a faster and more accurate query experience than using the keyword-based query; in the meanwhile, the workload of server and the network transmission pressure is also relieved. Furthermore, the developed syntactic and semantic feature extraction techniques help improve many images, audio and video analysis algorithms, which contribute to other related research areas. Finally, we have formulated a unified machine learning framework with a number of new machine learning algorithms for image and video contents classification and retrieval purposes. This unified framework and the associated algorithms establish a theoretical foundation in learning techniques, which can be widely applied to other research areas not limited to video content classifications.

The developed techniques are generally applicable to video-based education and entertainment. This project advances the state-of-the-art of digital video library research in several aspects: (1) the syntactic feature and semantic feature definition criteria and extraction algorithms, (2) the ways to fuse syntactic and semantic features, (3) the VCF classifiers that produce multi-layer video content class hierarchies, and (4) the category-based query mechanisms. The project also helps other application areas such as digital entertainment by engaging Augmented Reality to create new user experience.

2. Research Methodology and Achievements

This project develops various techniques to classify large-scale video stories effectively by engaging both syntactic features and semantic features in a unified learning framework. The original project proposal defines three research tasks: Task 1 – Syntactic feature and semantic feature definitions and extractions; Task 2 – Video content feature classification; and Task 3 – Experimental system implementation. A number of research results were achieved to enable content-based, personalized search techniques as well as to advance the state-of-the-art of digital video library research, multimedia information retrieval, and machine learning theoretical foundations in several aspects.

We describe our research achievements based on the specified major objectives in the following subsections.

2.1. Syntactic feature and semantic feature definition criteria and extraction algorithms

Our first piece of major work is multimedia feature extraction, i.e., text and visual feature representation. For text feature extraction, we have proposed a comprehensive and efficient scheme to detect, localize and extract text from real-world multilingual video sequences. For visual features, we have studied both low-level features and high-level semantic features from images and videos. Our extracted features have been successfully tested on a variety of multimedia applications, such as image classification and retrieval [1], video content summarization, etc.

2.2. Visual feature extraction from image sequences of video

One key work toward the video classification tasks is the video processing. In the past, we have already achieved a number of milestones in video processing domains. Specifically, we have developed a lot of video processing technologies, such as automatic segmentation of raw video sequences into individual video stories, commercial removal of news videos, etc. These technologies have been shown promising, in which some of them have been applied to existing commercial applications. The developed schemes in this objective are engaged in subsequent tasks of the project, facilitating the achievement of other research objectives as reported below.

2.3. Multi-layer video content class hierarchy mechanisms

Our major result in multi-layer video content classification is to attack the active learning problem in classification and retrieval tasks. Traditional approaches usually select only a single example in each learning round. Recently, we proposed the Batch Mode Active Learning (BMAL) technique, which is able to sample a batch of most informative examples in the active learning round. We formulated the problem by convex optimization and developed efficient algorithms to solve it. Our batch mode active learning techniques have been shown promising for both large-scale text classification tasks [2] and medical image classification problems [3].

We also established the multimodal and multilevel learning (MMML) framework toward video content classification and retrieval [4]. One of challenges in video classification and retrieval is how to fuse information from multiple modalities, such as textual and visual contents. To attack the problem, we suggested to model the structures of video data by graphical models and proposed a new semi-supervised learning scheme of fusing multimodal information over the graphs smoothly. Different from other heuristic approaches, our graph based multimodal fusion approach enjoys the merit of probabilistic interpretation from a random walk viewpoint.

Finally, to make our proposed learning framework practical for large-scale applications, we have suggested multilevel learning scheme of combining several learning methods of different computational costs and learning performances, including Nearest Neighbor (NN) Search, Support Vector Machines (SVM), and Semi-Supervised Learning, in a systematic solution. We have demonstrated promising results of our solution from empirical comparisons with other methods on large-scale news video datasets [4].

2.4. Unified learning framework for large-scale multimedia classification problems

We proposed a multi-layer content class hierarchy scheme together with a unified learning framework for large-scale video content classification tasks. In the multi-layer scheme, general binary classification can be naturally applied to form an effective solution for multi-class video classification purposes. To perform classifications, we have outlined a framework of unified machine learning framework intended to solve large-scale video classification tasks. Our unified learning paradigm (ULP) integrates Supervised Learning, Unsupervised Learning, Semi-Supervised Learning, Collaborative Learning, and Active Learning for solving the learning tasks effectively.

A significant achievement of our fundamental work is the newly proposed Unified Kernel Machine (UKM) solution [5]. Based on our previous unified learning framework, we proposed the UKM solution of unifying Supervised Kernel Machines, Semi-Supervised Kernel learning, and Active Learning techniques for tackling classification tasks comprehensively. The core component in the UKM is to develop a solution for learning effective kernels. We suggested the Spectral Kernel Learning (SKL) algorithm for learning kernels from both labeled and unlabeled data. The SKL algorithm is formulated by convex optimization and can be elegantly solved by Quadratic Programs, which performs better than traditional kernel learning approaches. We have applied the algorithms for data classification tasks and achieved very promising results.

2.5. A variety of machine learning algorithms for video content classification and retrieval

We further strengthened our investigations on more effective machine learning techniques with applications to the video content classification and search tasks. More specifically, we have recently worked out and achieved promising results in a number of learning tasks.

One critical task in our learning framework is to learn distance metrics or functions from contexts. To this purpose, we have proposed several novel algorithms for solving present challenges. For example, we formulated the regularized distance metric learning (RDML) technique [6] that can learn reliable metrics for tackling the noise problem in real-world applications. We also investigated techniques in learning kernel functions from users' query log, which is an important resource for video library. In particular, we proposed a novel technique to integrate the log information of user feedback into relevance feedback for image retrieval [7]. Our algorithm's construction was based on a coupled support vector machine which learns consistently with the two types of information: the low-level image content and the user feedback log. We presented a mathematical formulation of the problem and develop a practical algorithm to solve the problem effectively. We further proposed a unified framework for log-based relevance feedback that integrates the log of feedback data into the traditional relevance feedback schemes to learn effectively the correlation between low-level image features and high-level concepts [8]. Given the error-prone nature of log data, we presented a novel learning technique, named Soft Label Support Vector Machine, to tackle the noisy data problem. Experimental results showed that both of the above schemes were effective and promising. We also established a new framework of time-dependent query semantic similarity model on exploiting the temporal characteristics of historical click-through data [9], which can reflect real-world semantics such as real-world events that are happening over time.

Moreover, to reduce the human efforts in the learning tasks, we studied an important machine learning technique, Active Learning, intending to minimize the human labeling efforts in classification tasks. Our first solution is a Semi-Supervised Active Learning algorithm that can exploit information from both labeled and unlabeled data [10]. Experimental results on image retrieval applications have demonstrated that our technique clearly outperformed traditional methods. We also proposed two novel algorithms, the discriminative component analysis (DCA) and kernel discriminative component analysis (KDCA), for learning both linear and nonlinear metrics toward image retrieval and data clustering application [11]. Compared with other approaches, our algorithms are computationally efficient for large-scale applications.

In the theoretical learning foundation for image classification and retrieval, nonparametric models are often employed. The key challenge with nonparametric kernel learning, however, arises from the difficulty in linking the nonparametric kernels to the input patterns. We resolved this problem by introducing the graph Laplacian of the observed data as a regularizer when optimizing the kernel matrix with respect to the pairwise constraints [12]. We formulated the problem into Semi-Definite Programs (SDP), and proposed an efficient algorithm to solve the SDP problem. The extensive evaluation on clustering with pairwise constraints showed that the proposed nonparametric kernel learning method was more effective than other state-of-the-art kernel learning techniques.

2.6. Experimental system implementation and applications for video content classification and retrieval

In designing and implementing large-scale image repository for retrieval purpose with our techniques, we proposed a scalable content-based image retrieval scheme using locality-sensitive hashing (LSH), and conducted extensive evaluations on a large image testbed of a half million images [13]. To the best of our knowledge, there was less comprehensive study on large-scale CBIR evaluation with a half million images. Our empirical results showed that our proposed solution was able to scale for hundreds of thousands of images, which is promising for building web-scale CBIR systems. Furthermore, in applying our techniques to other applications, we investigated Augmented Reality area, which allows us to combine both strengths, improve existing game styles, and produce new games. Consequently, we implemented the *Augmented Reality Table* (ART), a prototype platform employing augmented reality technology to provide a virtual table for playing trading card games [14]. We illustrate ART as a system which engages augmented reality techniques to enable card games with more interactive and attractive user experience.

3. Summary and Future Research

In summary, we have identified the challenging issues for large-scale video content classification and retrieval problems, and proposed novel multi-layer video content class hierarchy mechanisms for tackling the research problems. To develop an effective solution, we have proposed a unified learning framework of integrating several state-of-the-art machine learning methodologies in a systematic fashion. Based on our framework, we proposed a series of innovative machine learning techniques for attacking the challenges successfully. A large-scale multimedia information retrieval testbed was implemented for experimental purpose, in which our proposed new algorithms can be encapsulated and evaluated. In our future work, we will continue to investigate further machine learning algorithms and apply these methods to large-scale video content classification and retrieval tasks.

4. Publications and Awards

[1] "CUHK at ImageCLEF 2005: Cross-Language and Cross-Media Image Retrieval," S. C. H. Hoi, J. Zhu and M.R. Lyu (2006). In *Proceedings of Cross Language Evaluation Forum (CLEF), Lecture Notes of Computer Science (LNCS 4022)*, pp.602-611, 2006.

[2] "Large-Scale Text Categorization by Batch Mode Active Learning," S. C.H. Hoi, R. Jin and M.R. Lyu, in *Proceedings 15th International World Wide Web conference (WWW2006)*, Edinburgh, UK, 2006.

[3] "Batch Mode Active Learning and Its application to Medical Image Classification," S. C. H. Hoi, R. Jin, J. Zhu and M.R. Lyu, in *Proceedings the 23rd International Conference on Machine Learning (ICML2006)*, Pittsburgh, Penn, US, June 25-29, 2006.

[4] "A Multimodal and Multilevel Ranking Framework for Content-Based Video Retrieval" S.C.H. Hoi and M.R. Lyu, in *Proceedings IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Special Session of Web Image and Video Search Technologies, 2007.

[5] "Learning the Unified Kernel Machines for Classification," S. C. H. Hoi, M.R. Lyu, and E. Y. Chang, in *Proceedings of The Twelfth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD2006)*, Philadelphia, USA, August 20 - 23, 2006.

[6] "Collaborative Image Retrieval via Regularized Metric Learning," L. Si, R. Jin, S. C.H. Hoi and M.R. Lyu, *ACM Multimedia Systems Journal*, pp.34-44, 2006.

[7] "Integrating User Feedback Log into Relevance Feedback by Coupled SVM for Content-Based Image Retrieval," S.C.H. Hoi, M.R. Lyu and R. Jin, in *Proceedings IEEE International EMMA Workshop in conjunction with 21st ICDE Conference*, Tokyo, Japan, April 2005, pp. 76-85.

[8] "A Unified Log-based Relevance Feedback Scheme for Image Retrieval," S. C.H. Hoi, M.R. Lyu, and R. Jin, *IEEE Trans. on Knowledge and Data Engineering*, vol. 18, no.4, pp.509-524, 2006.

[9] "Time-Dependent Semantic Similarity Measure of Queries Using Historical Click-Through Data", Q. Zhao, S. C. H. Hoi, T.-Y. Liu, S. S., Bhowmick, M.R. Lyu and W.-Y. Ma, in *Proceedings 15th International World Wide Web conference (WWW2006)*, Edinburgh, UK, 2006.

[10] "A Semi-Supervised Active Learning Framework for Image Retrieval," S.C.H. Hoi and M.R. Lyu, in *Proceedings IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2005)*, San Diego, CA, USA, June 20-25, 2005, pp. 302-309.

[11] "Learning Distance Functions from Contextual Constraints for Image Retrieval and Data Clustering", Steven C.H. Hoi, W. Liu, M. R. Lyu, Wei-Ying Ma, Proc. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR2006)*, New York, 17-22 June, 2006.

[12] "Learning Non-Parametric Kernel Matrices from Pairwise Constraints," S. C.H. Hoi, R. Jin and M.R. Lyu, in *Proceedings of the 24th Annual International Conference on Machine Learning (ICML 2007)*, OR, US, 20-24 June 2007.

[13] "An Empirical Study on Large-Scale Content-Based Image Retrieval", Y.-M. Wong, S. C.-H. Hoi, and M.R. Lyu, in *Proceedings of IEEE International Conference on Multimedia and Expo (ICME 2007)*, Beijing, China, 2-5 July 2007.

[14] "ART: Augmented Reality Table for Interactive Trading Card Game," A. H. T. Lam, K. C. H. Chow, E. H. H. Yau, and M.R. Lyu, *ACM International Conference on Virtual Reality Continuum and Its Applications (VRCIA 2006)*, Hong Kong, 14-17 June, 2006.

Automatic Multimedia Fission, Categorization and Fusion for Personalized Visualization in Multimedia Information Retrieval

Principal Investigator:

Professor Helen Meng [1]

Research Team Members:

Dr. Xie Lei, Postdoctoral Fellow [1]
(Currently Associate Professor at Northwestern Polytechnical University)
Mr. Shing-Kai Chan, MPh student [1]
Mr. Yuk-Chi Li, Research Associate [1]
Mr. Chuen-Fai Tse, Research Assistant [1]
Mr. Chuan Liu, Undergraduate student [1]
(Currently Research Engineer in the Chinese Academy of Sciences – The Chinese University of Hong Kong Institute of Advanced Integration Technologies in Shenzhen)
Mr. Xi Tan, Undergraduate Research Assistant [2]
(Currently Masters student at Cambridge University)

[1] *Dept. of Systems Engineering and Engineering Management, CUHK*
[2] *Tsinghua-CUHK Joint Research Center in Shenzhen, China*

Project Start Date: 1st March 2005
Completion Date: 31st December 2007

Abstract

Multimedia information portals in our global information infrastructure are hosting ever-increasing amounts of content in textual, graphical, audio and video forms. This creates a dire need for automatic technologies that enable organization and categorization of such massive repositories of information to facilitate browsing and retrieval by the user. This project explores three main aspects of multimedia processing: (1) Multimedia fission, which involves the use of acoustic, prosodic and lexical information for automatic story segmentation of a continuous stream of broadcast news. (2) Multimedia categorization, which uses a high-dimensional front-end feature vector together with pattern classification techniques for automatic classification of silence/speech/music segments in audio. (3) Multimedia fusion, which involves augmenting speech recognition transcripts of audio sound tracks with optical character recognition transcripts of video frame captions. In addition, we have also investigated the use of audio speech to drive video synthesis of an animated talking head. This report offers a brief description of various methodologies and techniques that have been developed and provides empirical performance results.

1. Project Objectives and Significance

Multimedia information portals in our global information infrastructure are hosting ever-increasing amounts of content in textual, graphical, audio and video forms. This creates a dire need for automatic technologies that enable organization and categorization of such massive repositories of information to facilitate browsing and retrieval by the user. In order to achieve this long-term goal, we address three main research problems: (i) multimedia fission – segmentation of continuous audio/video streams into self-coherent semantic units or constituents; (ii) multimedia categorization – categorization of segments of a continuous media stream based on content similarity; and (iii) multimedia fusion – combination of cross-media or cross-modality segments to fulfill the information needs of the user. The long-term significance of technology development in multimedia fission, categorization and fusion is to facilitate on-demand, just-in-time extraction and delivery of relevant information based on the needs of the user.

2. Research Methodology

On multimedia fission – we investigate possible methodologies for automatic story segmentation of Chinese broadcast news. Unlike textual documents that are clearly segmented via punctuation and typographic cues for sentences, paragraphs and sections; spoken documents take the form of a continuous stream of audio or video sound tracks that are in need of alternate cues for story segmentation. We investigate the use of segmentation cues such as pause durations, prosodic features and lexical patterns. We also develop statistical cue models and segmentation algorithms and conduct comparative empirical evaluation among alternative approaches for automatic story segmentation.

On multimedia categorization – we research approaches for audio scene analysis, in which a continuous audio stream is analyzed and classified into segments of silence, speech and music. These are the main audio scenes in Chinese broadcast news. We investigate the use of appropriate front-end features in combination with classification techniques.

On multimedia fusion – we investigate the use of optical character recognition to transcribe video captions into textual information, which is used to augment the audio transcripts generated by automatic speech recognition. The textual information derived from either modality is understandably imperfect, but should complement each other in facilitating information search by the user.

3. Results

3.1. Multimedia Fission

We investigated the use of audio *pauses, prosodic features* and *lexical patterns* for automatic story segmentation. Our experiments are based on the Voice of America Mandarin news broadcasts. Our first step utilizes audio pauses as an indicator for news story boundaries, while distinguishing them from other pause segments caused by breath groups, spoken language disfluencies, commercial breaks, etc. Our second step proposes the use of "*pitch reset*" for story segmentation. Pitch reset refers to the speaker's general behavior in pitch declination from high to low through the course of an intonational unit, followed by a reset to a high pitch at the start of the next intonational unit. A pitch reset is thus indicative of story boundaries. However, pitch reset is influenced by the inherent voice characteristics and speaking style(s) of the speaker. In addition, the tonality of the Chinese language presents an additional complexity in the use of pitch reset for story segmentation, when compared with non-tonal languages. For example, if the previous intonational unit ends on a high tone and the following intonational unit begins on a low tone, then the observed amplitude for pitch reset tends to be small. We also need to discriminate the use of pitch reset in utterance boundaries versus story boundaries. Analysis shows that story boundaries cannot be clearly discriminated from utterance boundaries by speaker-normalized pitch reset due to its large variations across different syllable tone pairs. Instead, speaker- and tone-normalized pitch reset can provide a clear separation between utterance and story boundaries. Experiments using decision trees for story boundary detection reinforce that raw and speaker-normalized pitch resets are not effective for Mandarin Chinese story segmentation.

Speaker- and tone-normalized pitch reset is a good story boundary indicator. Baseline results using only pause durations on single-speaker audio programs gave an *F*-measure of 82.2% for automatic story segmentation. Extensions with raw, speaker-normalized, as well as speaker- and tone-normalized pitch reset parameters gave *F*-measures of 68.3%m 77.4% and 86.7% respectively. The performance improvement from 82.2% to 86.7% is statistically significant. Analysis of the results also uncovered five major heuristics that show how speakers jointly utilize pauses, energy and pitch to signify segmentations in speech. The results are published (Xie, Liu & Meng, 2007).

We extend our work to use lexical patterns for automatic story segmentation. We devised a novel framework for *automatic extraction of named entities* from the speech recognition transcripts of radio audio. Named entities refer to names of persons, locations, organizations and numeric / temporal experiments. These are major information constituents in the audio content. Of particular interest is that our approach can handle *transliterated* Chinese names, e.g. 贸易代表巴尔舍夫斯基 (translation: trade representative Barshefsky). We link up repeated terms (i.e. words or word phrases) in the speech recognition transcripts to form *"lexical chains"*. It may be conceivable that lexical chain connections should be denser within a story but sparser across story boundaries. This is because crossing a story boundary brings about a topic shift which often manifests itself as lexical changes. More specifically, a high concentration point of starting lexical chains and ending lexical chains is considered as a hypothesis of a story boundary. It should be noted that the use of lexical chains for story segmentation is complicated by the fact that speech recognition transcripts are imperfect and contain errors. Furthermore, the Chinese language does not have an explicit word delimiter and there are ambiguities in word tokenization. We propose the use of the *log-normal model* to capture the statistical behavior of lexical chains, together with data-driven parameter selection for lexical chain formation. We compared segmentation performance with a suite of other statistical models, as well as a re-implementation of an alternative approach based on lexical chains as proposed by (Stokes 2004). Results indicate that the log-normal model achieves the best segmentation performance overall, where the F-measures improved from 46.8% to 64.1% based on the Voice of America Mandarin Speech Corpus (Li, Lo & Meng, 2006) (Chan, Xie & Meng, 2007).

3.2. Multimedia Categorization

We developed an audio scene analysis system to categorize segments of a continuous audio stream based on content similarity. The main audio scenes in the VOA Chinese broadcast news corpus are: silence, speech and music segments. Speech with background music is considered as speech segments as they carry spoken content that are indexed by automatic speech recognition. Silence segments may be identified by thresholding on short-time energy. Speech versus music discrimination is achieved by the use of a high-dimensional feature vector that includes LPCC (linear predictive cepstrum coefficients), LPS (line spectral pairs), MFCC (mel-frequency cepstral coefficients) and STFT (short-time Fourier transform) coefficients, totaling 94 in all. We also experimented with three classifiers that have demonstrated favorable performance in previous work, including KNN (K-nearest neighbor), MLP (multi-layer perceptron) and SVM (support vector machines). Experiments based on the Voice of America Mandarin news broadcasts show high classification performance with an F-measure of 98%. The SVM also strikes the best balance in terms of classification performance and computation time (real-time) among the three classifiers. This audio scene analysis system has been demonstrated at the China Hi-Tech Fair, Shenzhen Convention and Exhibition Center, October 10-18. 2007. The work has also been published in (Liu, Xie & Meng, 2007).

3.3. Multimedia Fusion

3.3.1. Augmenting automatic speech recognition transcripts of audio sound tracks optical character recognition transcripts of video frame captions

In multimedia retrieval of video information, the sound track can be indexed by automatic speech recognition, whose syllable accuracies may vary greatly, e.g. 27% to 60% for Hong Kong's TVB Cantonese broadcasts (Hui, Lo & Meng 2003). In addition, video captions can be indexed by optical character recognition (OCR). While recognition in either modality is understandably

imperfect, the complementarity across the two modalities implies that they may be used together to facilitate information search by the user. We conducted an initial investigation based on Hong Kong's TVB Cantonese news broadcasts.[1] Conventional OCR technology cannot be directly applied to video captions because these have a high degree of variability, as illustrated in Figure 1A to 1F. The illustrations above show some notable features of typical video frames from Hong Kong TVB Cantonese broadcast news. Frame (A) has a good contrast between the caption and the background. Frame (B) has a lower contrast due to similar colors between the caption and the background. Frame (C) has a complex background near the caption region. Frame (D) is a typical illustration of a white background. Frames like (C) and (D) generally have lower accuracies of caption recognition. There are also other frames without captions and need to be filtered, e.g. frames (E) and (F).

Figure 1 (A) Good contrast

Figure 1 (B) Insufficient contrast

Figure 1 (C) Complex background

Figure 1 (D) White background

Figure 1 (E) No caption

Figure 1 (F) No caption (commercial)

Figure1. Different types of captions found in Hong Kong TVB Cantonese Broadcast News.

[1] We have collected an audio-video corpus from the local television station – Hong Kong Television Broadcasting Ltd. Jade News, which is in Cantonese. The corpus consists of over 300 hours of video and about 50 have been transcribed. Such kinds of corpora of local relevance are lacking in the research arena. Our corpus should be a valuable resource for further research.

We developed methods for caption detection, enhancement, extraction and recognition. For caption detection, we utilized heuristics such as the fixed area in the video frame in which captions may appear and the dark lining of the conventional Chinese characters. We observe that captions in a video clip usually last for a sequence of frames in order to gain be perceptually registered by the human eye. Hence we apply a search algorithm to detect the appearance/ disappearance of video captions, which incorporates multi-frame integration techniques such as frame averaging and minimum pixel search. This reduces the complexity of the background. For caption enhancement, we applied color masking (i.e. thresholding on pixels because captions are monochromatic but the background has color), change masking (i.e. computing the difference between neighboring frames because background pixels tend to change more significantly than caption pixels) and region masking (i.e. detecting the area of regions because caption characters typically occupy hold regions with fixed sizes while the background typically occupies large blocks of connected regions). Figure 2A to 2C illustrate the outputs of various masking procedures:

Figure 2 (A). Effect of color masking, with the input on the left and the output on the right.

Figure 2(B) Effect of change masking – with (a) and (b) being neighboring frames and (c) being the output frame.

Figure 2(C) Effect of region masking, with (a) being the original frame (b) is the processed caption.

Caption character extraction utilizes the heuristic that the caption characters generally have fixed sizes and fixed gaps in between adjacent characters. Caption character recognition uses the commercial OCR software TH-OCR.™ Pilot experiments were conducted with a thirty-minute TVB news program from our collected corpus. Caption detection located 723 out of 735 video frames (98.4% accuracy). Caption enhancement with extraction located 4,375 out of 4,618 characters (about 94.7% accuracy). Caption recognition gave an accuracy of 91.3% with caption enhancement procedures, compared with 11.9% without. Figure 3 illustrates the output of caption recognition.

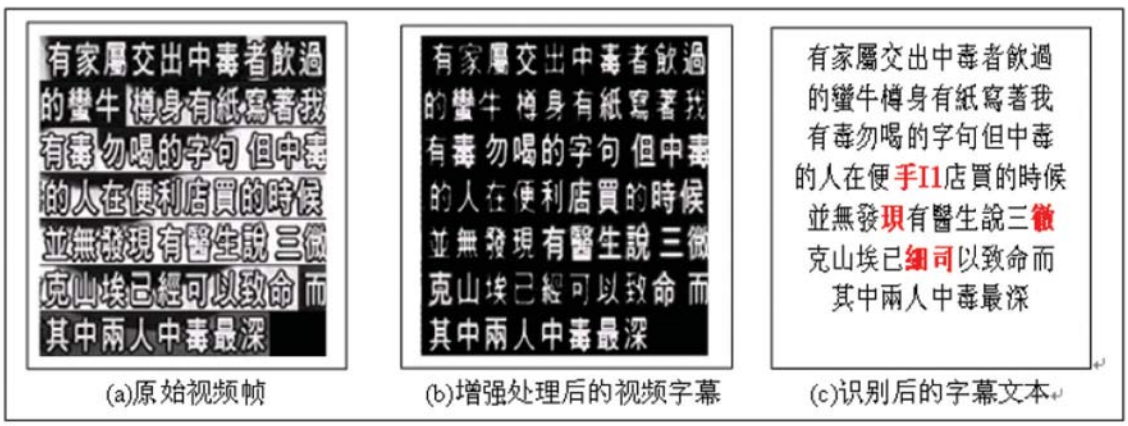


Figure 3. Illustration of the process of video caption recognition – (a) shows the original extracted captions from a video frame, (b) shows the enhanced captions and (c) shows the recognition outputs.

In addition to the above attempt to combine automatic speech recognition transcripts of audio sound tracks with optical character recognition transcripts of video frame captions, we have also another kind of information fusion, as described in the following subsection.

3.3.2. Speech-Driven Video-realistic Talking Face Synthesis

This is preliminary work that fuses audio and video by the use of an audio speech signal to drive the video synthesis of a talking head. This work leverages on previous work in speech-driven talking face synthesis for English using viseme (i.e. visual phonemes) based on (Xie & Liu, 2006). We have extended this approach to enable Cantonese speech-driven talking face synthesis. This involves input Cantonese speech which is transcribed by a Cantonese speech recognizer into a syllable sequence. We developed a translingual mapping scheme that performs symbolic mapping from Cantonese subsyllable units to English phonemes. An EM-based algorithm is then used to generate English visemes (or mouth animation parameters) that are synchronized with the raw Cantonese speech input. The synthetic talking face is used to augment the Cantonese speech input in Cantonese speech recognition experiments under noisy conditions and results show that the synthetic visual augmentation contributes to robustness in recognition performance. Another potential application for such technology is to provide a synthetic visual augmentation of the speech signal to facilitate human speech perception/communication under situations of hearing impairments that may be physical or environmental. This work has been reported in a conference paper (Xie, Meng & Liu, 2006).

References:

Hui, P. Y., Lo, W. K., Meng, H., "Two Robust Methods for Cantonese Spoken Document Retrieval", *Proceedings of the ISCA Workshop on Multilingual Spoken Document Retrieval (MSDR)*, Hong Kong SAR, China, April 2003

Stokes, N., Applications of Lexical Cohesion Analysis in the Topic Detection and Tracking Domain, PhD thesis, University College Dublin, 2004.

Xie, L. and Liu, Z.Q., "An Articulatory Approach to Video-realistic Mouth Animation," *Proc. ICASSP*, 2006.

4. Publications

[1] Chan, S.K., Xie, L. and Meng, H., "Modeling the Statistical Behavior of Lexical Chains to Capture Word Cohesiveness for Automatic Story Segmentation," *Proceedings of Interspeech*, 2007.

[2] Li, D., Lo, W.K. and Meng, H., "Initial Experiments on Automatic Story Segmentation in Chinese Spoken Documents using Lexical Cohesion of Extracted Named Entities", *Lecture Notes in Computer Science*, Vol. 4274, pp.693-703, 2006.

[3] Liu, C., Xie, L. and Meng, H., "Classification of Music and Speech in Mandarin News Broadcasts," *Proceedings of the National Conference on Man-Machine Speech Communication*, 2007.

[4] Xie, L., Meng, H. and Liu, Z.Q., "A Cantonese Speech-Driven Talking Face using Translingual Audio-to-Visual Conversion," *Lecture Notes in Computer Science*, Vol. 4274, pp.627-639, 2006.

[5] Xie, L., Liu, C. and Meng, H., "Combined Use of Speaker- and Tone-Normalized Pitch Reset with Pause Duration for Automatic Story Segmentation in Mandarin Broadcast News," *Proceedings of the North American Association for Computational Linguistics, Human Language Technologies Conference*, 2007.

[6] Xie, L., Liu, C. and Meng, H., "Discovering Salient Prosodic Cues for Automatic Story Segmentation for Mandarin Chinese Broadcast News," *journal manuscript*, forthcoming.

Automatic Video Segmentation and Tracking for Real Time Multimedia Services

Principal Investigator:

Professor King-Ngi Ngan [1]

Co-investigator:

Dr. Hongliang Li [1]

Research Team Members:

Mr. Qiang Liu, Research Assistant [1]

[1] *Dept. of Electronic Engineering, CUHK*

Project Start Date: 1st June 2006

Abstract

As an important technology, automatic video segmentation and tracking has great potential in a large number of application areas, including video monitoring and surveillance, video summarization and retrieval, video conferencing and videotelephony, computer vision and digital entertainment, etc. The purpose of this project is to develop techniques that are efficient and accurate in extracting interesting objects from videos in an unsupervised manner. Three key techniques, namely saliency model for video segmentation, saliency model for video tracking and boundary refinement technique, will be implemented. These include the use of perceptual features that will be modeled by several low-level and high-level cues for the object of interest. The outcomes of the project find applications in real time content-based multimedia services.

1. Significance

In recent years, there has been rapidly growing interest in content-based functionalities of video data, such as video editing, content-based image retrieval, video indexing, video event analysis, etc. Video segmentation as a challenging and active research area is a key technique for semantic object extraction and plays an important role in digital video processing and computer vision. The task of segmenting/tracking a video object emerges in many applications, such as traffic monitoring, surveillance and video conferencing, etc. Typically, the applications of video segmentation can be classified as follows:

- Video surveillance, where the segmentation result is used to allow the identification of an intruder or of an anomalous situation.
- Content-based video summarization, such as sports event summary, video skimming, video pattern mining, which requires the segmented semantic objects to perform the content classification.
- Content-based coding application.
- Computer vision, such as video matting, video toning, and rendering.

- Videotelephony and videoconferencing, where the segmentation can achieve better coding quality for the most relevant objects or to be able to store a specific object in a database, e.g., a face in a videotelephony directory.
- Digital entertainment, where some specific objects can be replaced by segmentation, such as the video games.

2. Objectives

- To investigate the saliency model for extracting objects of interest from videos.
- To develop novel automatic video object segmentation techniques for generic and specific (e.g., video surveillance, videoconferencing) applications based on saliency model.
- To develop real time object segmentation techniques based on saliency model and fast transform.

The successful completion of this project will provide new video segmentation tools for many applications in multimedia services such as videotelephony, videoconferencing, computer games and digital entertainment.

3. Research Methodology

Generally, object segmentation can be divided into two stages, i.e., desired object detection and object extraction, which are concerned with pattern recognition and clustering techniques, respectively. According to the detection mode, object segmentation can be performed in two manners, i.e., supervised or unsupervised. The emphasis of this proposed research is to develop techniques that are efficient and accurate in extracting interesting objects from video sequences in an unsupervised manner. The major contribution of this project is on the use of perceptual saliency map that will be modeled by several low-level and high-level cues for the object of interest (OOI). These cues include color, orientation, intensity, location, and motion, etc. Fig. 1 shows the framework of our implementation approach.

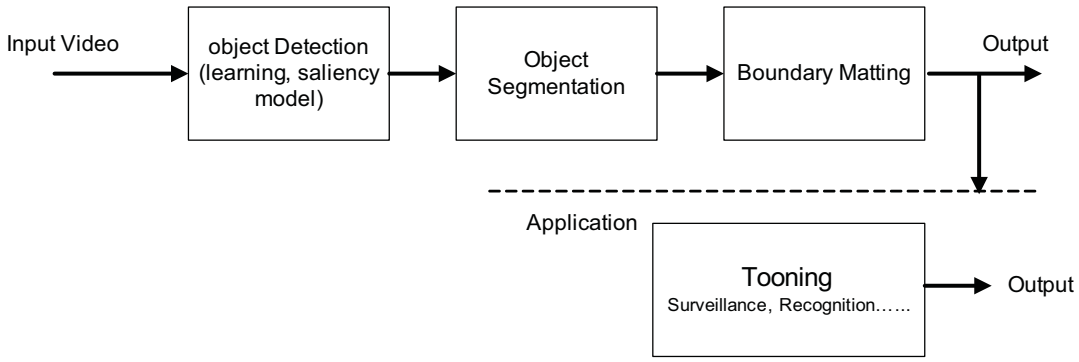


Fig. 1. The framework of the developed object segmentation system.

For example, for the attention-based object segmentation method, we first construct the saliency map using the inherent features of the OOI. The general expression can be modeled by

$$S = f(oc, oi, op, os)$$

where S denotes the value of saliency map, and, *oc*, *oi*, *op* and *os* correspond to the color, intensity, position and structure of the object of interest, respectively.

Then, we employ non-linear filtering to eliminate the noise in the obtained saliency map, and use the classification approach to extract the object regions. Certainly, in order to refine the boundary of the objects, several optimization methods may be used, such as maximum a posteriori probability.

4. Results Achieved So Far

A. Attention-based video segmentation: segmentation based on facial saliency map (Completed)

We first generate the saliency map from the input head-and-shoulder type video image by our proposed facial attention model. Then, a geometric model and an eye-map built from the chrominance components are employed to localize the face region according to the saliency map. The final step involves the adaptive boundary correction and the final face contour extraction. There are three conspicuity maps corresponding to the chrominance, luminance, and position information used to construct the facial saliency map, which can be used to locate the potential face areas. Experimental evaluation on test sequences shows that the proposed method is capable of segmenting the face area effectively, which is evident in Fig. 2.

Original Image | Facial Saliency Map | Segmented Result

Fig. 2. Segmentation based on facial saliency model.

B. Attention-based video segmentation: segmentation of defocused video image (Completed)

In this work, an unsupervised segmentation algorithm based on matting model is proposed to extract the focused objects in the low depth of field video images. Based on the visual attention idea, we have constructed the focused saliency model to measure the focused region. Nonlinear filtering and matting approaches are employed to segment the object of interest accurately. An example can be found in Fig.3.

Fig. 3. Segmentation result for defocused video image.
Left: The original image. Right: Segmentation result based on the proposed saliency model.

C. Object detection (Completed)

- **Feature Extraction: Fast and Efficient Method for Block Edge Classification (Completed)**

In this work, we investigate the difference properties from three coefficients in the non-normalized Haar transform (NHT) domain and present a fast and efficient method to classify block edge using these properties. The proposed method significantly reduces the number of computational operations in the edge models determination with no multiplications and less addition operations.

- **Object Detection based on extracted features (Completed)**

In order to extract objects of interest (human) from videos, we first perform the object detection to obtain the location of the object. Based on the extracted features, we use NHT coefficients and block edge information to detect the object by using cascade classifier. Firstly, we manually segment the over 12,000 face samples from a large number of face databases. Then, a cascade of face classifiers is trained to detect the frontal face. Here, we use AdaBoost learning approach to perform the training process, which can capture the face saliency features for future face detection. Currently, we have finished six layers training work. For each layer, over 50% non-face samples will be rejected, while falsely eliminating only 0.1~0.3% face samples. The proposed face detector scans the image at many scales, looking for face locations within the scaled windows. It consists of three parts, i.e., skin color filtering, rejector cascade, and cascades of boosted face classifier, which is shown in Fig. 4.

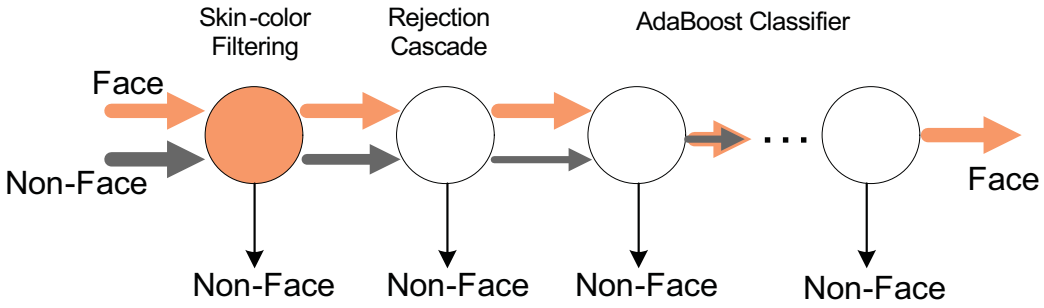


Fig. 4. The flowchart of the human face detection procedure.

D. Face segmentation based on graph-cut (Completed)

After the face detection, we can access the face locations in the input image rapidly. In this work, we use the minimal-cut optimization method to perform the face segmentation algorithm in the obtained face regions.

Generally, when the foreground or background is clearly defined by the user inputs, the segmentation process can be carried out based on graph cut optimization. Unfortunately, in the unsupervised manner, we only have the coarse object locations that are obtained from the boosting object detector. The corresponding sub-window may cover the complete object contour or only some parts of face regions. It means that the pixel outside the window may belong to the background, while the pixel inside the window is likely to be part of the face. We cannot determine with confidence which pixel should be marked as the background or foreground.

The proposed method works at two levels, i.e., coarse and fine scales. The initial segmentation is performed at the coarse level. There are some regions used for estimating foreground information, whilst four regions for the background. We take their means and variances as the initial clusters for the background, which means there are four components to model the GMM in the background. For each pixel in this region, we compute the weighted distances to the face and background clusters according to cost function and the similarity with other. Finally, we use the minimum cut to perform the global optimization.

The second level is the finer segmentation, which aims to refine the initial segmentation result. The corresponding foreground regions are defined as the set of the pixels belonging to the body terminal in the current window, while the background regions consist of those pixels outside the window that are classified to the background terminal. We use 8 and 8 components to describe

the foreground and background colors, respectively. The mean and covariance of component *k* are estimated based on the K-means algorithm. Then, the similar method as the coarse scale can be used to compute the energy *E*1 and *E*2. Note that in the data cost function *E*1, the weight of each component is estimated based on the spatial samples within a defined window of 20x20 pixels that is centered with the current node. The min-cut is also used for the final optimization. Some examples based on this method are shown in Fig. 5.

Fig. 5. Graph cut based segmentation result for human faces.

E. Real-time human body segmentation and matting (in progress)

In the future work, we aim to develop an automatic algorithm to segment human body from real-time video, which can be applied to videoconferencing and video coding.

5. Publications

[1] Hongliang Li, King N. Ngan, and Zhenyu Wei, "Fast and efficient method for block edge classification and its application in H.264/AVC video coding", *IEEE Transactions on Circuits and Systems for Video Technology*, accepted.

[2] H. Li, and K.N. Ngan, "Unsupervised video segmentation with low depth of field", *IEEE Transactions on Circuits and Systems for Video Technology*, in press.

[3] H. Li, and K.N. Ngan, "Automatic video segmentation and tracking for content-based applications", *IEEE Communications Magazine*, vol. 45, no. 1, January 2007, pp. 27-33.

[4] Q. Liu, C. Cai, K.N. Ngan and H. Li, "Camshift based real-time multiple faces match tracking", *International Symposium on Intelligent Signal Processing and Communication Systems*, China, Dec. 2007.

[5] H. Li and K.N. Ngan, "Unsupervised segmentation of defocused video based on matting model", *International Conference on Image Processing*, Atlanta, U.S.A., October 2006, pp. 1825-1828.

Information Retrieval from Mixed-language Spoken Documents

Principal Investigator:

Professor Tan Lee [1]

Co-investigator:

Professor P.C. Ching [1]

Research Team Members:

Dr. Wentao Gu, Research Associate [1]
Ms. Houwei Cao, PhD student [1]
Mr. Yu-Ting Yeung, MPh student [1]

[1] *Dept. of Electronic Engineering, CUHK*

Project Start Date: 1st June 2006

Abstract

This research aims to develop the key technologies for information retrieval from mixed-language spoken documents. With low-cost storage devices and rapid internet connection, we now have easy access to a hugeamount of audio and video information. Effective indexing and retrieval techniques are needed to make this information useful to our daily life and work. Existing audio indexing systems assume that the recording involves only one language with known identity. They are not able to handle mixed-language speech, which is very common nowadays. Hong Kong is an international city where many people are Cantonese and English bilinguals. English words are frequently embedded into spoken Cantonese. This presents great challenges to automatic transcription and indexing. This research addresses the problems of acoustic-phonetic variation, pronunciation and lexical variation in mixed-language speech. A speaker-independent speech recognition system is developed for automatic transcription of mixed-language recordings from radio and TV programmes, lectures, meetings and interviews. A prototype system for efficient information retrieval will also be implemented.

1. Objectives and Significance

With low-cost storage devices and rapid internet connection, we now have easy access to an unmanageable amount of multimedia information. Since speech is the most natural and convenient way of information exchange for general people, audio recordings of human speech are among the most commonly available resources. Examples include radio and TV programmes, lectures, meetings, and interviews. Recently online video/audio sharing, e.g., YouTube, has become increasingly popular. This leads to an explosive growth of public accessible audio information with virtually unrestricted content. Effective indexing and retrieval techniques are needed to make the information useful to our daily life and work. Hong Kong is an international city where many people, especially the young generation, are Cantonese and English bilinguals. English words are frequently embedded into spoken Cantonese, e.g. "能夠同佢哋 work together 我覺得好 exciting." For information retrieval from audio archives that contain mixed-language content, conventional approaches based on monolingual speech recognition are not applicable, simply because there is no prior knowledge about when language switching would occur.

This research project aims to develop the key technologies for information retrieval from mixed-language spoken documents. A speaker-independent continuous speech recognition system will be developed for automatic transcription of Cantonese-English code-mixing speech. Based on the analysis of a large amount of real speech data, this system is designed to deal with the pronunciation, lexical and acoustic variations that are caused specifically by code-mixing. The speech transcription generated by the system contains multiple hypotheses to facilitate flexible and admissible indexing. At the end of this project, we plan to implement a prototype system for efficient information retrieval from mixed-language recordings.

2. Research Methodology

2.1. Understanding the Problem

The phenomenon that English words are frequently inserted into spoken Cantonese sentences is referred to by linguists as *code-mixing*. Cantonese is known as the primary language or matrix language while English is the secondary or embedded language. In a typical Cantonese-English code-mixing sentence, a Chinese word is substituted by its English equivalent. The grammatical structure is totally based on Cantonese. Since the speaker is usually a native speaker of the matrix language, it is inevitable that the embedded English word carries Cantonese accent to certain extent. The Cantonese accent may be realized in many different ways. For example, the syllable structure of an English word may be modified to become acceptable in Cantonese. Phonemes that are unique to English may be replaced by Cantonese phones that sound similar. Thus English words in code-mixing utterances should not be treated as the same as in a monolingual utterance from a native English speaker. On the other hand, code-mixing happens less frequently in read-style speech than in casual conversational speech. In casual speech, speakers may not strictly follow the pronunciations as specified in a standard dictionary. Pronunciation variation is expected to be the major problematic issue in code-mixing speech recognition.

2.2. Data Collection

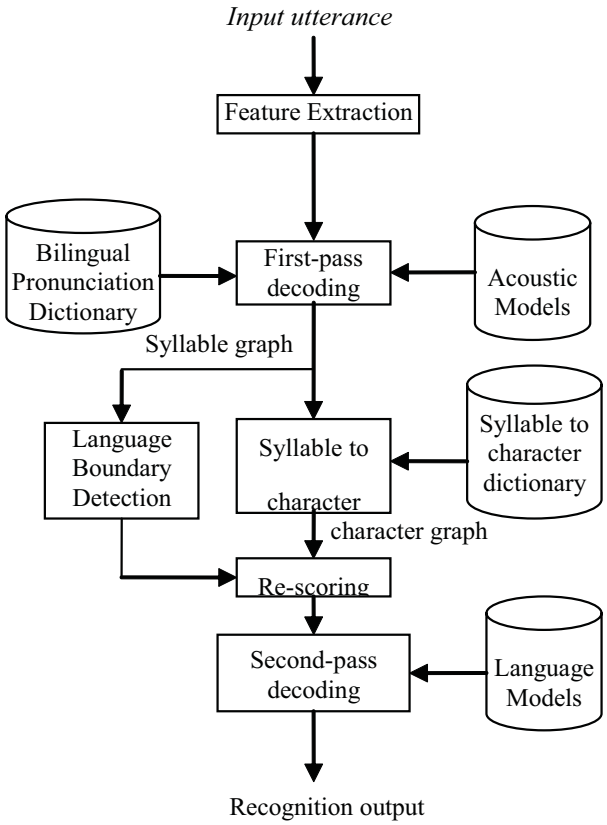
To facilitate pronunciation variation analysis and acoustic modeling, a large number of code-mixing utterances are needed. The design of code-mixing sentences is done manually based on previous linguistic studies, text materials from local newspapers and online resources. Most of the sentences contain one English segment. The embedded English words are those commonly used in code-mixing speech and they cover different word categories and word lengths. There are practical difficulties in collecting a large amount of text data to facilitate statistical language modeling of code-mixing speech. Cantonese is a spoken dialect. It does not have a standard written form that is on par with standard written Chinese. Written Cantonese is neither taught in schools nor recommended for official and documentary usage. On the other hand, code-mixing is a domain-specific phenomenon. It is found in the discourses that involve contemporary and cross-cultural issues, e.g., computer, business, fashion, food, entertainment and showbiz. In our study, Cantonese text data are selected from three major sources, namely newspaper, magazines; and online diaries. Preliminary manual inspection is done to identify the sections or columns that are highly likely to contain code-mixing text, and subsequently articles that contain Cantonese-specific characters are selected.

2.3. System Design

The major components of a state-of-the-art continuous speech recognition system are acoustic models, a pronunciation lexicon, and language models. The acoustic models are a set of hidden Markov models (HMMs) that characterize the statistical variation of the input speech features. Each HMM represents a specific sub-word unit, e.g., a phoneme. The pronunciation lexicon and language models are used to define and constrain the way in which the sub-word units can be concatenated to form words and sentences. The training of acoustic models and language models requires a large amount of speech and text data. The pronunciation lexicon is often adapted from linguistic resources.

The following figure shows the design of a code-mixing speech recognition system. We need a set of cross-lingual acoustic models that cover both English and Cantonese. The pronunciation lexicon and the language models also have to include both Chinese and English word entries. The

recognition process is divided into two passes. In the first pass, the search space is significantly reduced from an all-included grammar network to a syllable/word graph. The basic elements of the graph are nodes and arcs. Each arc represents a hypothesized Cantonese syllable or a hypothesized English word. In the second pass, the syllable graph is further converted into a character graph from which the most likely code-mixing sentence is decoded. For robust indexing, we may use the syllable graph directly, which provides more hypotheses than a single top-best sentence.



3. Progress and Results

3.1. Corpus Development

3.1.1. Speech corpus

We have completed the development of CUMIX, the first Cantonese-English code-mixing speech corpus. It was designed with the same standard as the commonly used monolingual speech databases for speaker-independent speech recognition. CUMIX contains not only code-mixing utterances but also monolingual English words, which are useful for the study of accented English. The content of these utterances were designed based on text materials from local newspapers/ magazines, online diaries, and newsgroups. A total of 19000 utterances were recorded from 40 male and 40 female native Cantonese speakers. Most of the code-mixing utterances contain one English segment, which is either a single word or a sequence of words that occurs frequently in daily speech, e.g., “around”, “concern”, “CD”. All utterances have been manually transcribed. In addition, a small amount of spontaneous speech data with unpredictable degree of code-mixing has also been collected. The recordings include two-party free discussions on some selected topics, meetings, classroom tutorials and presentations.

3.1.2. Text corpus

We have compiled a code-mixing text corpus with approximately 6.8 million Chinese characters. The text data were extracted from three major sources, namely newspaper, magazines; and online diaries. There are about 4,600 unique Chinese characters in the database.

3.2. Pronunciation Variation Analysis

A comprehensive analysis of pronunciation variation in code-mixing speech has been carried out. It covers both the matrix language and the embedded language. We use three speech corpora, namely, TIMIT – native English, CUSENT – read-style Cantonese, and CUMIX – Cantonese-English code-mixing speech. Monolingual English and Cantonese acoustic models are trained with TIMIT and CUSENT respectively. These models are used to recognize test utterances from matched or mismatched sources. First, under the matched condition, i.e., test utterances are from the same corpus as the training data, major pronunciation variants are identified from the most confusable phoneme pairs. They form the baseline for subsequent analysis. When test utterances from mismatched sources are used, the recognition accuracy drops significantly. Comparing the respective phoneme confusion patterns with the baseline, additional pronunciation variants caused by code mixing can be found.

The results of our analysis confirm that English words spoken by Cantonese speakers indeed carry strong Cantonese accent. Ten related pronunciation variations are identified, e.g., "tutor" → "tuta", "bit" → "beat", "three" → "free". It is also observed that Cantonese speakers delete consonant syllable codas in most cases. Comparing casual with read-style Cantonese speech, the same patterns of pronunciation variation are observed. However, the degree of variation is much greater in code-mixing speech than in read speech. These findings are very useful to the design of a cross-lingual phoneme inventory for acoustic modelling.

3.3. Cross-lingual Acoustic Modeling

Acoustic models in a speech recognition system map the acoustic features of speech signals to different sound units in the language. In the context of Cantonese-English code-mixing, a fundamental question to be answered here is whether we should have two sets of language-dependent phoneme models or a cross-lingual phoneme set that covers both Cantonese and English. Based on the pronunciation variation analysis, cross-lingual modeling approach is considered more appropriate. Although Cantonese and English are phonetically and phonologically very different, they share some common phonemes. These phonemes can share the same acoustic models in the cross-lingual modeling, so that better utilization of the training data can be achieved. We design the cross-lingual phoneme set by including all Cantonese phonemes and those English phonemes that have no Cantonese equivalents. The total number of phonemes is 70. The acoustic models are context-dependent hidden Markov models. Each model consists of three emitting states and each state is represented by a mixture of 32 Gaussian components. The training data are from CUSENT and CUMIX. The word-level recognition accuracy is about 60% for the code-mixing test utterances in CUMIX.

3.4. Language Modeling

The major problem in language modeling for code-mixing speech recognition is due to the limited amount of data. The English words in our text corpus generally have limited number of occurrences, which is not sufficient for probability estimation. To address this problem, we adopt a class-based approach. The English words are divided into 15 classes according to their parts of speech (POS) or meanings. It is noted that many of the words are nouns. Thus finer classification of nouns is implemented. The language models are evaluated in the phonetic-to-text (PTT) conversion task. Assuming that the true phoneme transcription is known, the language models are used to determine the most likely character/word sequence that matches the transcription, like in an LVCSR system. We use the syllable transcriptions of the CM test utterances of CUMIX as the input. A conversion accuracy of 91.5% has been attained.

In our previous research, a standard Chinese text corpus with 98 million characters was compiled. We use a rule-based translation approach to convert this corpus into a spoken-Cantonese text corpus. The rules are extracted from on a set of manually translated sentences. By increasing the training data, it is expected that the performance of language models can be much improved, especially for the Cantonese part.

3.5. Language Boundary Detection

There is no prior knowledge about when there is a switch of language in a code-mixing utterance. Language boundary information, if available, is very useful to code-mixing speech recognition. Our approach makes use of the syllable/word graph generated by the first-pass decoding. English words generally have longer duration than Cantonese syllables since they may contain multiple syllables. Under the assumption that each utterance has one English segment, the English word with the longest duration in the lattice is most likely to indicate a correct recognition result, and the start and end time of the word are taken as the language boundaries. This results in a detection accuracy of 82.3% for the code-mixing utterances in CUMIX.

3.6. Prosody Analysis of Code-Mixing Speech

Prosody is an important component of speech. It refers to the rhythm, tempo and intonation of a continuous utterance. We believe that prosody-related features, namely pitch and duration, are useful to automatic recognition of code-mixing speech. We start by analyzing monolingual Cantonese utterances and investigate the effect of focus and tonal context. It is observed that a Cantonese syllable under emphatic focus tends to have a longer duration and wider pitch range than one without focus. The duration lengthening and pitch range expansion extend to the vicinity of the focused syllable. We have also collected a small amount of code-mixing speech data for prosody analysis and the work is on-going.

4. Publications and Awards

[1] Joyce Y.C. Chan, P.C. Ching, Tan Lee and Houwei Cao, "Automatic speech recognition of Cantonese-English code-mixing utterances," In *Proc. INTERSPEECH 2006*, pp.113 – 116, Pittsburgh, USA, September 2006.

[2] Houwei Cao, P.C. Ching, Tan Lee and Ning Wang, "An Extended Cantonese-English Code-mixing Speech Corpus: exCUMIX," In *Proc. Oriental COCOSDA*, Penang, Malaysia, December 2006.

[3] Wentao Gu and Tan Lee, "Effects of tonal context and focus on Cantonese F0," In *Proc. International Congress of Phonetic Sciences*, pp.1033-1036, Saarbrücken, Germany, August 2007.

[4] Wentao Gu and Tan Lee, "Effects of focus on prosody of Cantonese speech - a comparison of surface feature analysis and model-based analysis," in *Proc. ParaLing'07*, pp.59-64, Saarbrücken, Germany, August 2007.

[5] Houwei Cao, P.C. Ching and Tan Lee, "Pronunciation variation analysis for Cantonese-English code-mixing speech," to be presented in *Oriental COCOSDA*, December 2007.

[6] Joyce Y.C. Chan, Houwei Cao, P.C. Ching and Tan Lee, "Automatic recognition of Cantonese-English code-mixing speech," under review by *IEEE Transactions on Audio, Speech and Language Processing*.

Wireless Mesh Network Testbed

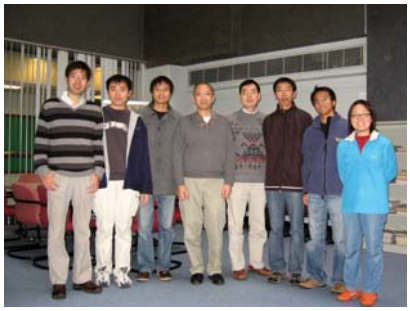
Principal Investigator:

Professor Dah-Ming Chiu [1]

Co-investigators:

Professor Wing-Cheong Lau [1]
Professor John C.S. Lui [2]

[1] *Dept. of Information Engineering, CUHK*
[2] *Dept. of Computer Science and Engineering, CUHK*



Project Start Date: 1st June 2006

Abstract

Wireless mesh networks is a hot research topic. The algorithms for routing and resource allocation for wireless mesh networks are very challenging problems because it is very difficult to precisely model the physical layer fading and interference exactly. Therefore a wireless mesh network for experimentation is very useful. The main goal of this project is to build a wireless mesh network using commodity hardware and public domain software components as much as possible. The usefulness of the wireless mesh will be tested by experimentation by a wireless network routing research project (using path aggregation).

1. Introduction

Most research of wireless networks nowadays rely on simulation. However, simulation is unable to capture the complex signal irregularity in wireless channel, like in an indoor environment. This motivates the building of a wireless mesh network Testbed. The objective of this project is to setup a wireless mesh network testbed in the two engineering buildings by deploying about 30-40 mesh routers in different floors and locations. Various research experiments can then be performed in this network.

In the rest of this report, we report mainly on the progress of setting up the wireless mesh network, including: Hardware components, Software components, Architecture and deployment scheme, Management system, and Overall appraisal. In each case, we report on the work completed and on-going work.

After a report on building the wireless mesh network, we also briefly describe a research project “Path aggregation to support VoIP in wireless mesh networks”, which will try to use the mesh network for experimentation.

2. Hardware Components

2.1. Work Completed

Two types of routers are used in the testbed: the Pepwave Manga Dual PCI unit and the Linksys WRT54GL.

Pepwave Manga Dual PCI unit

Linksys WRT54G

The main component of the testbed is the Pepwave Manga Dual PCI unit. Although it is more expensive than its Linksys counterpart, it has two mini-PCI slots which allows the installation of two 802.11a/b/g wireless interfaces. With multiple interfaces per node, multiple frequency channels can be assigned to a router. This reduces the interference between mesh nodes and provides a platform for more potential research projects, like resources planning. The presence of 802.11a mode in the Atheros mini-PCI Wi-Fi cards in Pepwave router also allows us to completely free from the interference from numerous existing 802.11b/g devices in the engineering building. 30 Pepwave routers have been purchased. They form the backbone of the mesh testbed.

Although with only one single interface, the Linksys WRT54GL provides a low cost component for our testbed. It can be placed at the edge of the network to offer last mile wireless coverage. The number of wireless interface of Linksys router can be further increased by attaching a wireless access point to the LAN of the router. It has been tested that the wireless bridging function of WAP54G and DWL-7100AP allow us to construct point-to-point wireless links between routers. The limitation of this configuration is that these wireless links cannot form any loop. As a result, extra caution and manual configuration are needed.

The following table lists the two routers used in this project

	Pepwave Mange Dual PCI	Linksys WRT54GL
Cost	HKD$1550	HKD$430
Machine Architecture	ARM	MIPSEL
CPU	ARM 822Tid at 166MHz	Broadcom 5352 at 200MHz
Memory	RAM: 32MB/ Flash:16MB	RAM: 16MB/ Flash: 4MB
Radio	802.11a/b/g x 2	802.11b/g x 1

2.2. On going work

The USB port in the Pepwave router allows us to further expand its functionality. For example, we can connect it with an external hard disk to increase its storage capacity. By connecting an extra USB Wi-Fi card, the number of interfaces per node can be increased.

3. Software Components

3.1. Work Completed

As the software components run in embedded system have different machine architecture with PCs and require cross compilation before they can be used, we have modified and tested several open source software components of the mesh routers. This includes:

1. Embedded Linux: each router is loaded with an embedded Linux, an operating system that specifically run on embedded devices, such as router. The embedded Linux provides a platform for the running of other programs, e.g. the OLSRD routing daemon.

2. OLSRD: It is an implementation of the Optimized Link State Routing protocol. With the routing daemon installed, the routers can discover each other and compute the best path between any two nodes. There are several reasons of choosing OLSRD as our routing daemon. Firstly, as all of the routers in the testbed are stationary, OLSR which uses a proactive approach to discover route and maintain routing information can reduce route setup delay. OLSRD also has the gateway function, which allows a mesh router to act as a gateway to the Internet. Finally, it supports the use of multiple interfaces and multiple subnets, which increases the scalability of the network. It has been verified that OLSRD has no known problem to run in both Pepwave and Linksys routers by cross compiling it with the corresponding cross complier. Routing daemons in different system can communicate with each other without any problem too.

3. iperf: in addition, we also port iperf to the routers. Iperf is a network utility which allows the user to measure the TCP and UDP bandwidth and performance. It is useful in network planning and research experiments.

4. Management system daemon: we have implemented a daemon used by the management system which collects information from the router and receive instructions from the server.

3.2. On going work

More tools for performance measurement and management can be installed in the routers. One ongoing work is to cross compile and test these applications so that it can be run correctly in these devices.

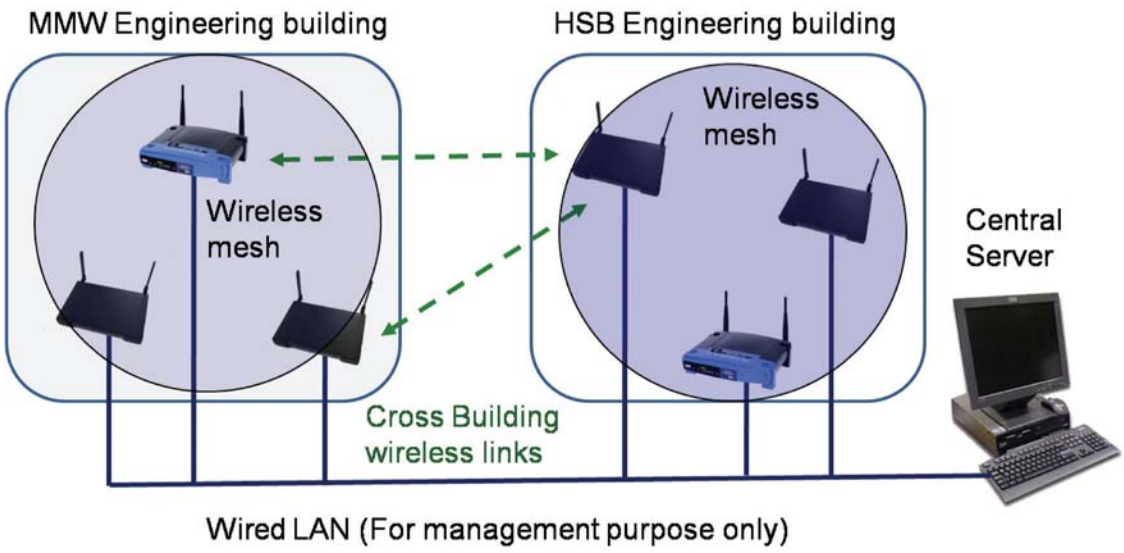
The Project title (on the first page) should be centered, completely capitalized, and in Times 14-point, boldface type. The PI's and Co-Is' name(s) appear below the title in capital and lower case letters. Reports with multiple Co-Is may require two or more lines for this information.

4. Architecture and Deployment Scheme

4.1. Work Completed

20 Pepwave and 5 Linksys routers are currently deployed in the two engineering buildings. They are connected to each other with the wireless interfaces, forming a wireless mesh network. Additionally, each router is also connected to the central server through a wired LAN. The wired connections are for management purpose only and may not be present in typical wireless mesh network. As the objective of the wireless mesh testbed is to carry out various experiments, the wired connections allow the server to remote control, upload/download files and collect experimental data from the routers.

A few wireless links have also been setup to connect the two engineering buildings. Although these cross building links are long, their performances are good. It is because there is direct Line-Of-Sight between the end points of these links. The diagram below shows the architecture of the testbed:



4.2. On going work

A few routers are currently reserved for testing purpose. We are going to expand the mesh by deploying them to cover some dead spots. The network will also be extended to the Computer Science and Engineering department. A challenge will be connecting and managing the two networks together as the LAN of the two departments is separated.

5. Management System

5.1. Work Completed

Configuring a large number of routers is a time consuming and tedious task. To manage the routers in a systematic manner, the patch system and management system are developed.

1. Patch system

A patch script is installed in each router. As the server and routers are connected by wired connection, from the central server, we can do the following with a single command:

a. Upload files to or download files from the routers.
b. Configure the interfaces of the routers, such as IP addresses and subnet.
c. Execute commands in a batch automatically.

2. Wireless network management system

A wireless network monitoring system is also developed to manage and visualize the information of the routers. The system contains the following components:

a. Client daemons. They are installed in all routers. The client daemons collect interface, routing information from the routers and receive command from the server.
b. A central server. It retrieves the information from the routers, stores information into a database and receives commands from the user.
c. A database which keeps the most updated information of all routers.
d. A web-based graphical user interface. The user interface displays the network topology and nodes' information. It can also accept input from the user and provide advance functions like disable link and route tracing.

Advantages of this architecture:

a. We separate management system in the server side into three components: the server, the database and user interface. A clear separation of the components allows us to update a part of the system more easily without affecting the functions of the other parts. For example, we can change the display format without modifying the server program.

b. We adopt the web based graphical user interface in our system. This allows us to have access of the interface from other computers through LAN or the Internet.

5.2. On going work

We may further optimize and stabilize the system. For example, the server uses HTTP connections to retrieve data from the routers. However, the overhead of this operation is large. Additionally, more functions such as displaying the link qualities of the network can be added to the management system.

6. Overall Appraisal of the Wireless Mesh Construction

About two third of the routers are deployed and running in the two engineering buildings. In the next step, the network will be further expanded with the remaining nodes. More functions will be added to the test-bed to provide more tools for various research activities.

7. Experiments for the “Path Aggregation” Project

The “path aggregation for VoIP in wireless mesh” project concerns how to route VoIP calls so that multiple VoIP flows can share the same path, and merge multiple VoIP payloads into the same physical packets. Path aggregation can potentially produce significant savings in bandwidth of a wireless mesh network, This is the research topic of Zhengyi Zhang’s MPh thesis. In the ensuing months, Mr Zhang plans to use the wireless mesh network technology developed in this project to performance experiments to validate his research ideas.

Real-time Transmission of High Definition (HD) 3D Video and HD Audio in Gigabit-LAN

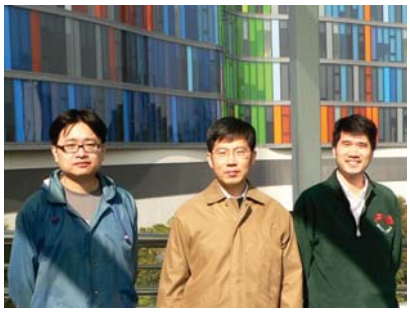
Principal Investigator:

Professor Raymond Yeung [1]

Co-investigators:

Mr. Alan Lam [1]
Mr. Ka-Kui Choy [1]

[1] *Dept. of Information Engineering, CUHK*



Project Start Date: 1st June 2007

Abstract

With the advances in electronic appliances and gigabit networks, we believe that real-time transmission of HD 3D (stereoscopic) video will be of high demand and will prevail in the next generation Internet, especially in the areas of medicine, product design, and digital entertainment. However, with the current multimedia technologies, we are still facing lots of challenges in HD 3D video real-time transmission. This project will focus on the difficulties of the deployment of HD 3D video real-time transmission and try to work out state-of-the-art solution to overcome these difficulties.

Why HD, 3D, and Real-time Transmission

As digital TV will be available as early as 2007/08, more and more HD products are released in the market. It is projected that HDTV telecasting will soon become a demand in HK, in a way very much like color TV replacing black-and-white TV in the 1970s. Real-time HD transmission technologies will become a need within few years.

Under the trend of globalization, collaboration among different regions will be commonplace. Sharing 3D images and videos among partners at different remote sites for such purposes as architecture review, manufacture design, or remote medical diagnosis, is desired. Furthermore, 3D display will also be a key component in future digital entertainment.

The medical field has high demand on 3D HD video transmission for remote surgery or diagnosis. However, due to the bandwidth and equipment limitation, it can currently be done only in the 2D SD format. Robotic surgical systems are also becoming popular because of its precision and minimum invasion. To operate such a system, the surgeon needs to have a clear and 3D view. And the 3D video transmission needs to be in real-time because a split-second delay during a blood vessel cut can be fatal in any operation. Real-time 3D video transmission also plays an important role in future network game industry.

Existing Problems

1. **High bandwidth of HD video**

   For 8bits 4:2:2 1080i HD Video, the video bandwidth for uncompressed video is around 1.69Gbps. Although gigabit connection is widely available today, it still cannot fulfill the requirement for transmitting 2D HD Video.

2. **No solution for HD video real-time transmission**

   Data compression by MPEG-2 TS or H.264 can help to solve the HD video high bandwidth problem. However, they need high-end server support (i.e., very expensive hardware card, high processing power CPU, and large amount of RAM) which is not available in conventional PCs or electronic appliances. In addition, these video coding formats hardly address the problem of real-time transmission.

3. **Lack of effective real-time transmission protocol for Audio Engineering Society (AES) or multi-channel audio**

   Most AES are for streaming only. So far, there is no real-time transmission protocol for any AES or multi-channel audio.

4. **Lack of video CODEC for HD 3D video**

   There exists no video CODEC for 3D video or real-time HD video transmission. Without an effective video CODEC for HD 3D video, real-time HD 3D video transmission will be extremely expensive.

5. **Auto-stereoscopic display on heterogeneous devices**

   More and more electronic appliances can now display Auto-Stereoscopic 3D images, such as 3D LCD monitor or cellular phone. However, there is no effective algorithm or standard to display a single source 3D image on these heterogeneous devices.

Project Objectives

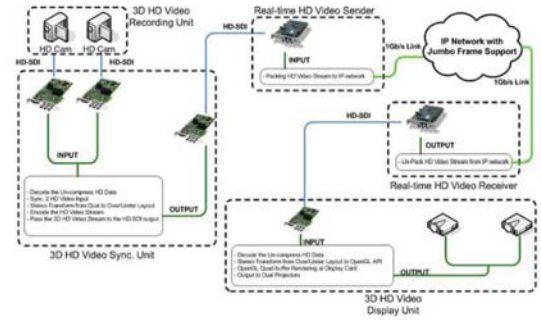
In this project, new standards and protocols will be developed to address the problems in real-time transmission of HD 3D video.

a. Develop a new CODEC for capturing and trans-coding the 3D HD Cameras.
b. Develop a new video format standard that accommodates 3D video and HD audio.
c. Develop a new compression algorithm that strikes a balance between compression rate and coding complicity.
d. Develop a new transmission protocol for 3D HD video.
e. Develop a new 3D display algorithm for Auto-Stereoscopic display on heterogeneous devices.

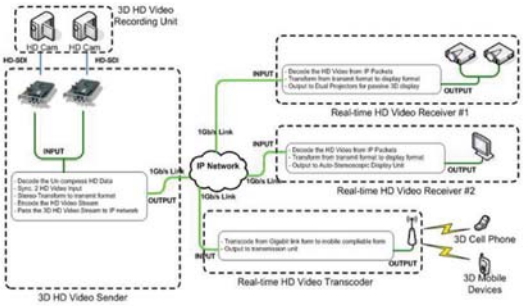
Status and Progress

Prototype and Developing Platform

The following prototype has been built to demonstrate the feasibility of real-time 3D HD video transmission on the computer network and have been demonstrated to the public on CUHK Orientation Day. By evaluating the prototype, we found that some lag and delay have been generated by the internal process, the hardware and the network protocol stack. In order to further reduce the delay and minimize the lag, we plan to move our development platform from Windows to a Linux-based platform.



Block diagram of the Prototype



Block diagram of the developing platform

Collaboration with Department of Surgery, CUHK

3D video of minimally invasive surgical operation has been captured from DaVinci Robotic Surgical System and has been successfully telecasted on the computer network in real-time.

DaVinci Robotic Surgical system

Real-time 3D video transmission

This project has supported the International Robotic Surgery Symposium 2007 organized by CUHK Surgery Department. Surgical operation 3D video was telecasted in real-time on the computer network during live demonstration sessions. The real-time 3D video transmission from Robotic Surgical System "Da Vinci" in this Symposium is a showcase to demonstrate such an application of real-time HD 3D video transmission and distribution for next-generation telemedicine for biomedical research, training, and education as well as health-care. With further advances in high-speed broadband optical access networks worldwide, such HD 3D video transmission and distribution over very long distances may make this an even more useful system for tele-medicine across the continents. This project has also helped HK Trade Development Council to demo 3D Minimally Invasive Surgery Video in Medical and HealthCare Fair 2007. We have also given a seminar and demo of "3D HD Video Distribution for Robotic Surgery" in a HKIE and SPIE activity. The photos of these events can be viewed at http://www.ine.cuhk.edu.hk/l2/web/shiae/album/index.html

Conducting 3D Video Conference with Tsinghua University on IPv6

3D video conferences have been conducted with Tsinghua University on IPv6 and Internet2 network after Tsinghua has duplicated our 3D display units. Now we are fine-tuning the 3D video conference setting and preparing the HD 3D video conference trial test via gigabit link.

3D Video Conference Setup

High Dynamic Range Image Compression and Display

Principal Investigator:

Professor Leo Jiaya Jia [1]

Research Team Members:

Dr. Wei Feng, Postdoctoral Fellow [1]
Mr. Qi Shan, Research Assistant [1]

[1] *Dept. of Computer Science and Engineering, CUHK*

Project Start Date: 1st June 2007

Abstract

The high dynamic range (HDR) images are the new formats of the multimedia data which contain rich and broad visual information in colors and structures. Only in the recent a few years, with the development of the image and video acquisition devices, the HDR data attract much attention due to their faithful representation of the ubiquitous high contrast scenes in the real world. In this project, we plan to explore the challenging problems of efficiently and robustly rendering, compressing, and visualizing the HDR images.

1. Objectives and Significance

The research on high dynamic range (HDR) images compression has been made significant progress in recent years. However, there are still many processing and visualization problems unsolved. For instance, current LCD or CRT monitors have the global contrast no more than 1000:1. They are not capable of showing the visual details of the HDR images in which the contrast may exceed 10^5 : 1 [1]. One example is shown in Figure 1 that the HDR images contain rich structural information. When displaying the HDR image on ordinary displays, details are lost in saturated pixels.

Figure 1: The HDR image displayed on ordinary monitor under different exposures loses details in either bright or dark regions.

Therefore, the high dynamic range should be compressed for the purpose of visualization. Another problem is related to the limitations of the image and video encoding methods, such as MPEG and JPEG, in representing the HDR format. These widely used media data compression standards mostly support 8-bit color channels. Without proper compression of the large-size HDR data, the adoption of this new format in many possible applications is hindered.

The objective of this project involves exploring the challenging problems of efficiently and robustly compressing and visualizing the HDR images. The investigation will be carried out in both the theoretical study and the empirical validation of the novel processing algorithms and visualization devices. The long-term goal of this project includes:

- To create an area of strength on the research of processing HDR data, and push the frontier of the HDR study in medical imaging and 3-D rendering.
- To build the applicable HDR data acquisition and testing environment to facilitate the algorithm benchmark and validation on hardware.
- To establish a regional platform of researching the innovative HDR applications on a variety of joint areas, such as the image enhancement.

2. Research Methodology

In this project, we plan to accomplish our objectives by conducting several tasks, specifically, in HDR data acquisition, compression, display, processing, and encoding respectively.

HDR data acquisition. The HDR images are usually taken by the high-end cameras using “raw” format or constructed by calibrating a set of low dynamic range images with fixed positions under varied exposures. However, it may not be feasible or easy for a user to setup the HDR image acquisition environment especially when the scene has very high dynamic range or contains moving objects. Thus, we will investigate the method to realistically synthesize the HDR images from ordinary low dynamic range(LDR) images [2]. The specific energy minimization problem based on the features of the LDR images can be solved to improve the visual quality of the result.

HDR image compression. Normally, the acquired HDR images cannot be easily visualized using the ordinary display or encoded using the common bmp or jpeg image format [3,4,5]. The process to reduce the high dynamic range while preserving the visual details is called range compression. In order to produce a visually satisfying compression result, on one hand, the local contrast should be maintained in a perceivable level; on the other hand, the global contrast of the HDR images should be primarily reduced. To satisfy these requirements, in this project, we shall study the window-based tone mapping method in which a global optimization problem [6] can be solved with appropriately guided local linear constraints. **Specifically, our method operates on the windows containing only a few neighboring pixels. In each window, we shall use a linear function to constrain the modification of the radiance values in a monotonic order to retain the local structures. Our proposed method does not involve scale decomposition, layer separation, or image segmentation, thereby is immune from the possible artifacts such as noticeable seams between segments.**

HDR data processing. Although there have been many algorithms proposed to calibrate the images, till now, there is still no many methods dedicated to processing the HDR data. It is noted that in photo-editing tools, processing the HDR images and demonstrating the desirable effect done on them is more important than just viewing them with sufficient visual details in dark and bright regions. The possible processing may include denoising, sharpening, stylization, or smoothing. So it is essential for the HDR data, before and after these operations, to not only reveal all structure details, but also faithfully illustrate the modifications performed. We thus propose to process the HDR data visually faithful to the compression.

Interactive touch-up. It is not rare that the user desires a direct image touch-up to locally modify the luminance or contrast in the processing results. We, thus, plan to provide flexible interaction tools [7] in our system using the simple brush strokes to collect pixel samples and to indicate the modifications. The process is not isolated on these samples. It fits into our global optimization framework to produce seamless image and video result.

3. References

[1] Ahmet Oguz Akyuz, Roland Fleming, Bernhard E. Riecke, Erik Reinhard, and Heinrich H. Bulthoff. “Do hdr displays support ldr content? a psychophysical evaluation” In *SIGGRAPH*, page 38, 2007.

[2] F. Banterle, P. Ledda, K. Debattista, and A. Chalmers. “Inverse tone mapping” In *GRAPHITE*, pages 349–356, 2006.

[3] K. Devlin. “A review of tone reproduction techniques” Technical Report CSTR-02-005, 2002.

[4] Durand F. and Dorsey J. “Fast bilateral filtering for the display of high-dynamic-range images” In *ACM SIGGRAPH*, pages 257–266, 2002.

[5] Raanan Fattal, Dani Lischinski, and Michael Werman. “Gradient domain high dynamic range compression” In *ACM SIGGRAPH*, pages 249–256, 2002.

[6] William W. Hager, Shu-Jen Huang, Panos M. Pardalos, and Oleg A. Prokopyev. *“Multiscale Optimization Methods and Applications (Nonconvex Optimization and Its Applications)”* Springer, 2005.

[7] Berthold K. P. Horn. “Determining lightness from an image” *Computer Graphics and Image Processing*, 3:277–299, 1974.

Multimedia Content Distribution over Hybrid Satellite-terrestrial Communication Networks

Principal Investigator:

Professor Jack Y.B. Lee [1]

Co-investigator:

Professor Peter T.S. Yum [1]

[1] *Dept. of Information Engineering, CUHK*

Project Start Date: 1st June 2007

Abstract

The Japan Aerospace Exploration Agency (JAXA) together with the Japan National Institute of Information and Communications Technology are planning to launch a new communication satellite called the Wideband Internetworking Engineering Test and Demonstration Satellite (WINDS) in *Feb 2008*. The WINDS satellite will provide high-speed data communication for the Asia Pacific countries including Hong Kong with a goal to promote the use of satellites in such fields as Internet communications, education, medicine, disaster measures and Intelligent Transport Systems. This research proposal aims at leveraging the access to the WINDS satellite for the research and experiment of multimedia content multicasting over the WINDS satellite system. This project has two objectives. The first objective is to explore new research opportunities and to tackle the research challenges in delivering high-bandwidth multimedia contents such as audio and video to a large number of users across multiple countries and regions. The second objective is to offer a content distribution platform for JAXA WINDS members to multicast research, cultural, and educational contents to other JAXA WINDS members as well as ordinary Internet users.

1. Introduction

The Japan Aerospace Exploration Agency (JAXA) together with the Japan National Institute of Information and Communications Technology are planning to launch a new communication satellite called the Wideband Internetworking Engineering Test and Demonstration Satellite (WINDS) in Feb 2008. The WINDS satellite will provide high-speed data communication for the Asia Pacific countries including Hong Kong with a goal to promote the use of satellites in such fields as Internet communications, education, medicine, disaster measures and Intelligent Transport Systems.

Over the past several years we have been participating in the international collaboration efforts organized by JAXA and with the endorsement of JAXA we have submitted in July 2006 a formal application proposal to the Japan Ministry of Internal Affairs (MIC) to request access to and conduct experiments on the WINDS satellite. With the support from JAXA our application had been approved in May 2007.

This research proposal aims at leveraging the access to the WINDS satellite for the research and experiment of multimedia content multicasting over the WINDS satellite system. This project has two objectives. The first objective is to explore new research opportunities and to tackle the research challenges in delivering high-bandwidth multimedia contents such as audio and video to a large number of users across multiple countries and regions. The second objective is to offer a content distribution platform for JAXA WINDS members to multicast research, cultural, and educational contents to other JAXA WINDS members as well as ordinary Internet users.

2. Research and Development

The current Internet does not yet support multicast transmission over a global scale. Therefore distributing high-quality multimedia contents to a large number of users is still very costly. By contrast, satellite communication is inherently broadcast in nature and thus could offer a cost effective solution for multicasting high-quality multimedia contents. For regions not covered directly by the satellite, we could extend the network's reach by routing and forwarding data over the terrestrial Internet. As shown in Figure 1 the media server sends out data over both the satellite and the terrestrial Internet.

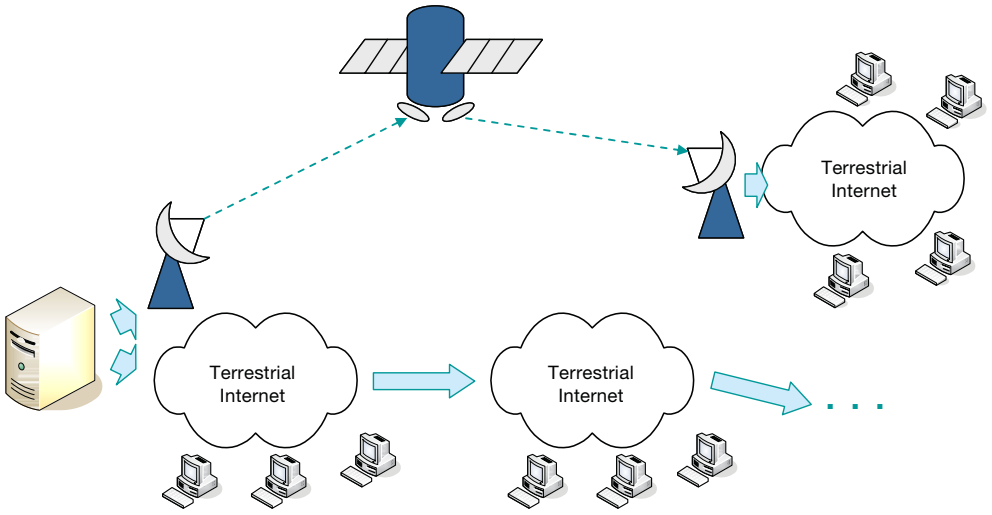


Figure 1 – Hybrid Satellite-Terrestrial Media Multicast

Given there are now two network channels available, one research challenge is on the routing of media data to a user, who may have access to either one or both of the network channels. Beginning with the simplest routing policy, the server can simply be configured to route media data over the satellite channel whenever the user is known to have satellite connectivity (e.g., identification via user ID or IP address). Otherwise the media data will be routed via the Internet using application-layer multicast (ALM).

Taking this idea one step further, the system can also exploit the fact that some users will have access to both network channels, and thus instead of receiving data through a fixed network channel, can monitor the network performance and then dynamically route the data through the best-performing network channel. This idea can be extended even further by exploring the transmission of media data over both Internet and satellite channels simultaneously – multi-path routing. This will be more complex as the characteristics of the Internet channel and the satellite channel, in terms of bandwidth, delay, and loss, are very different. Therefore new media streaming algorithms will need to be developed to compensate for the differences and to exploit the benefits of multi-path routing.

3. Existing Strengths

This research project comprises two components, namely the network communications module and the content management module. The former is the focus of this research project while the latter will be supported by the CUHK Video Multicast Platform already developed and deployed for use in the Internet. With the CUHK Video Multicast Platform, this project will extend it with a satellite communications module to access the JAXA WINDS satellite for multicasting media contents to JAXA WINDS members. This will enable the system to stream media data to three types of users, namely unicast-only, multicast-ready, and satellite-ready users as shown in Figure 2.

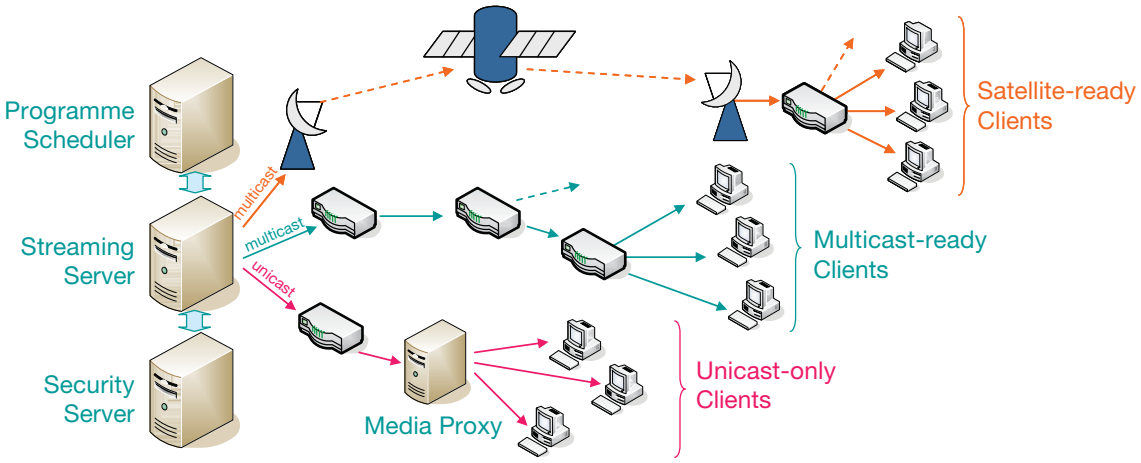


Figure 2 – System architecture of the multimedia content distribution platform

In terms of deployment, there are two possible models. The first one is a semi-centralized model where the primary servers are hosted at CUHK, which JAXA WINDS members can upload their contents for distribution. The second one is a distributed model where individual JAXA WINDS members setup their own CUHK Video Multicast Platform for direct content distribution. The wider deployment of the CUHK Video Multicast Platform will enable us to collect experimental data that will be very useful for the research community.

4. Summary

The JAXA WINDS project is a unique and rare opportunity to conduct networking and multimedia research over a real satellite environment. The choice of using multimedia content distribution as a platform for research in this project offers two advantages. First, distributing high-bandwidth multimedia content over both wired, wireless, and satellite communication links is a unique problem that is not well studied. Therefore we expect to uncover new research problems that may open up new directions in satellite communication research. Second, the CUHK Video Multicast platform, if used or adopted by other WINDS members inside and outside Japan, will enhance the impact of this research project and potentially generate new international research collaborations in the future.

Shun Hing Distinguished Lecture Series

To achieve the Institute's mission to promote appreciation of engineering in society through education programs, the Institute has organized a Shun Hing Distinguished Lecture Series. So far, **ELEVEN** distinguished lectures have been presented by renowned scholars. These lectures all were very well received and we will continue to line up and invite outstanding researchers to visit CUHK and to deliver distinguished lectures for the Institute.

Bioengineering: The Engineering Discipline of the 21st Century?

by Professor Yongmin Kim

Professor and Chair
Department of Bioengineering
University of Washington, USA

Date: 18th October 2005, Tuesday

Abstract

Bioengineering is the fastest growing engineering discipline in the U.S. The areas of research and education/training covered in Bioengineering are vast: from biomedical imaging to bioinformatics and from therapeutic devices to molecular bioengineering and tissue engineering. Unique integration of biology, engineering, nanotechnology, information technology and medicine has been underway in Bioengineering, continuously creating new frontiers and opportunities in both research and education. Also, Bioengineering has taken (and will continue to expand) a key role in translation of biomedical science and technology research into the clinic and/or home to enable improvements in prevention, diagnosis and therapy, ultimately transforming the delivery of health care and enhancing the quality of our lives.

Computational Models for Medical Image Analysis

by Dr. Nicholas Ayache

Research Director
Epidaure/ Asclepios Laboratory, INRIA, France

Date: 18th October 2005, Tuesday

Abstract

Medical image analysis brings about a revolution to the medicine of the 21st century, introducing a collection of powerful new tools designed to better assist the clinical diagnosis and to model, simulate, and guide more efficiently the patient's therapy. A new discipline has emerged in computer science, closely related to others like computer vision, computer graphics, artificial intelligence and robotics.

In this talk, I will describe the increasing role of computational models of anatomy and physiology to guide the interpretation of complex series of medical images, and illustrate my presentation with three applications: the modeling and analysis of 1) brain variability from a large database of cerebral images, 2) tumor growth in the brain and 3) heart function from a combined exploitation of cardiac images and electrophysiology.

I will conclude with a presentation of some promising trends, including the analysis of in vivo microscopic images.

Exploring Nonlinear Constraint Optimization and Their Applications

by Professor Benjamin W. Wah

Department of Electrical and Computer Engineering
And the Coordinated Science Laboratory
University of Illinois, Urbana-Champaign, USA

Date: 16th June 2006, Friday

Abstract

Constraint optimization exists in many engineering applications, including production planning, system design, artificial intelligence, control systems, semiconductor manufacturing, and operations research. A key observation on the constraints of many of these application problems is that they are highly structured and involve variables with strong spatial or temporal locality. This observation is reasonable because constraints are often derived based on variables related in time or in space. Based on this observation, large-scale application problems can be partitioned by their constraints into a small number of much simpler subproblems. Because each subproblem has only a fraction of the original constraints, it is a significant relaxation of the original problem and has an exponentially lower complexity. As a result, many problems that cannot be solved before can now be solved easily. In this talk, we illustrate this approach in solving some large-scale applications in spacecraft planning, VLSI cell placement and routing, satisfiability, neural networks, and operations research.

From Functional Molecules to Complex Systems

by Professor Chih-Ming Ho

Ben Rich – Lockheed Martin Professor
Director, Institute for Cell Mimetic Space Exploration
University of California, Los Angeles, USA

Date: 4th October 2006, Wednesday

Abstract

The rapid developments of nanotechnologies have driven the production of molecular-scale devices towards the functionalizing of materials, directly manipulating of genetic molecules and engineering strains of proteins to possess novel functionalities. The human body of meter size is an extremely complex adaptive system. The question of how we will span from the nano-scale molecules to complex systems with much large length scales and eventually will enable us to enrich human lives is not obvious, but a key task.

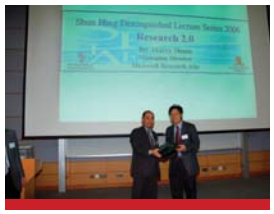
Cell fuses genetic informatics with functional molecules to result in the smallest complex system in nature. In this presentation, we will illustrate the capability of inducing cell toward desired destiny as an example of controlling a complex system across a wide span of length scales, from nanometer to microns. Part of the challenge is that millions of functional molecules inside a cell are governed by networks of signal pathways which are still beyond comprehension. In addition, while a cocktail of cytokines is used to yield cells toward a final desired phenotype, exploring all of the possible combinations of these stimuli can result in a very large number of trials. The test covering all the combinations can not be carried out in a realistic time frame to find the most potent combination. Applying the engineering feedback control concept, we can search and reach the optimum condition with a very small number of trials. We have demonstrated that optimally designed time-varying stimulations can self-organize and adjust the functionalities across multiple length scales to efficiently reach the desired state. This may yield new insight into unlocking and acquiring novel control modalities of the underlying mechanisms that drive the natural processes of life.

Research 2.0

by Dr. Harry Shum

Managing Director
Microsoft Research Asia, China

Date: 28th November 2006, Tuesday



Abstract

Two important technical trends have emerged over the last decade. First, the Web continues to grow its size with a variety of new data and penetrates every aspect of our lives. Second, "software as services" has become not only a new form of software delivery, but also a way of releasing quality software. These two trends are having sweeping influence on the software ecosystem and IT industry by driving many online businesses that connect people to people and people to information. In this talk, we examine the potential impacts of "Web" and "software as services" on "research." We share our findings with the audience by suggesting six new directions for "Research 2.0": (1) The Web as a research platform; (2) leveraging community effects; (3) data centric computing; (4) the need of deployment driven research; (5) infrastructure is crucial; and, (6) seamless user experience across devices. To illustrate these important directions of Research 2.0, we will describe and demonstrate some recent progress made at Microsoft Research Asia in these areas. We will also show demos of Chinese Couplets (智能中文對聯系統) and Video Messenger with special effects.

A quick overview will also be given of Microsoft Research Asia, "the world's hottest computer lab" (MIT Technology Review, June 2004), including activities in basic research, technology transfer, product incubation, technology licensing, university relations and internship program.

Pattern Recognition: Bayes or not Bayes, Is that the Question?

by Professor Fred Juang

Motorola Foundation Chair Professor and
Georgia Research Alliance Eminent Scholar
Digital Signal Processing Group and
Telecommunications Group
School of Electrical & Computer Engineering
Georgia Institute of Technology, USA

Date: 2nd December 2006, Saturday

Abstract

Pattern recognition is a well-studied problem which recently regains its significance in machine learning, artificial intelligence, data mining, and the general area of inferential computing. A theory developed by Thomas Bayes (1702-1761) on optimal decision has long served as the mathematical foundation of statistical pattern recognition techniques. In this talk, we review and re-examine this well-established framework in the context of practical problems and ask the question: Is Bayes minimum risk an achievable goal and if possible how to reach that goal? We analyze the question against backdrop of modern techniques and come to the formulation of a performance-based approach to the fundamental problem of pattern recognition that holds the potential of achieving Bayes' optimal result.

Optimum Space-Time Block Code Design for MIMO Communications: A Case for Linear MMSE Receivers

by **Professor Kon Max WONG**

NSERC-Mitel-CITO Professor of Signal Processing
Chair, Department of Electrical & Computer Engineering
McMaster University, Canada

Date: 6th December 2006, Wednesday

Abstract

The explosive expansion in wireless communications in recent years encounters severe technical challenges including the demand of transmitting signals at high rates under strict limitations of power and bandwidth in an environment rich of scattering. Multi-input multi-output (MIMO) wireless links are important recent developments in wireless communication systems due to their enormous potential in meeting these challenges. Existing MIMO communications employ M transmitter antennas and N receiver antennas and design space-time codes to enable the exploitation of both the high performance provided by the space diversity and the high data rate afforded by the capacity available in the MIMO channels. Recent research on space-time code design to exploit these factors has been mainly targeted for the MIMO system equipped with a maximum likelihood (ML) receiver since an ML receiver is able to achieve the full diversity gain provided by the multiple antenna system. However, ML receivers are non-linear receivers well-known to have very high complexity often prohibiting practical implementation. Thus, there is a need to study space-time code designs for MIMO systems employing linear receivers.

In this lecture, we examine the design of optimum space-time block codes (STBC) for MIMO systems equipped with a linear MMSE receiver. From the analysis of the probability of detection error of the system, we show the necessary and sufficient structures of a STBC for achieving minimum bit-error rate (BER) from which the diversity gain of such a system is then derived. Indeed, the diversity gain of such a system, even when employing optimum STBC, is shown to be inferior to that of a MIMO system equipped with an ML detector. We then propose that the optimum STBC design principle be extended to cover the transmission of multiple blocks of data. Examination of the BER for the multi-block transmission system shows us that the extended optimum STBC design has necessary and sufficient structures parallel to those of the optimum single-block design. Analyses reveal that increasing the number of blocks covered by the optimum code design increases the diversity gain, yet the order of the normalized detection complexity remains virtually constant. Theoretical predictions confirmed by computer simulations showed that the number of blocks covered by the optimum STBC design does not have to be very large for the system performance to catch up with and surpass that for an ML receiver. Utilizing this optimum code design extended to multi-block transmission, for applications permitting latency of signal reception, the MIMO wireless system can be brought to practical implementation without losing any of its advantages.

Novel One Dimensional Nanostructures

by Dr. M. Meyyappan

Chief Scientist for Exploration Technology
Center for Nanotechnology
NASA Ames Research Center, USA

Date: 15th January 2007, Monday


Abstract

The combination of remarkable mechanical properties and unique electronic properties of carbon nanotubes(CNTs) offers significant potential for revolutionary applications in electronics devices, computing and data storage technology, sensors, composites, storage of lithium for battery development, nanoelectromechanical systems(NEMS),and as tip in scanning probe microscopy(SPM) for imaging and nanolithography. Thus the CNT synthesis, characterization and applications touch upon all disciplines of science and engineering. This talk will provide on overview of CNT growth, and application development in many of the above areas. The ability to grow inorganic nanowires with controlled properties and vertical orientation provides another competitive avenue for some of the needs mentioned above. Our work in this direction will also be described.

Prosody Generation for Communicative Speech Synthesis

by Professor Yoshinori Sagisaka

Professor, Global Information & Telecommunication Institute

Language and Speech Science Research Laboratory
Waseda University, Japan

Date: 9th February 2007, Friday


Abstract

A corpus-based prosody modeling is proposed aiming at F0 control for communicative speech synthesis. For input information in communicative speech synthesis, input word attributes are employed. Two experimental results are introduced to show the possibility of communicative speech prosody control from input word attributes. The results showed that F0 height, F0 dynamic patterns and duration could be consistently controlled by the word attributes. The positive-negative characteristics can be controlled by F0 height, while confident-doubtful, allowable-unacceptable characteristics were reflected in F0 dynamic patterns and duration. Through these analyses, the correlations between perceptual impressions on output speech and input word attributes are turned to be useful for F0 characteristics prediction from input word attributes.

The Multimedia Communications Revolution of the 21st Century

by Professor Lawrence Rabiner

Center for Advanced Information Processing(CAIP)
Rutgers University, USA

Date: 28th June 2007, Thursday


Abstract

We are now in the midst of a Multimedia Communications Revolution in which virtually every aspect of telecom is changing in ways that would have been considered unthinkable just a decade or so ago. Perhaps the greatest challenge in realizing this communications revolution is to figure out how to provide a range of new services that seamlessly integrate text, sound, image, and video information and to do it in a way that preserves the ease-of-use and interactivity of conventional telephony, irrelevant of the bandwidth or means of access of the connection to the service. In order to achieve this overarching goal, there are a number of technological problems that must be considered, including:

- compression and coding of multimedia signals, including algorithmic issues, standards issues, and transmission issues;
- synthesis and recognition of multimedia signals, including speech, images, handwriting, and text;
- organization, storage, and retrieval of multimedia signals;
- access methods to the multimedia signal;
- searching;
- browsing.

In each of these areas a great deal of progress has been made in the past few years, driven in part by the relentless growth in processing and storage capacity of VLSI chips, and in part by the availability of broadband access to and from the home and to and from wireless connections.

It is the purpose of this talk to review the status of the technology in the areas of telecom, multimedia compression, and multimedia understanding, and to illustrate some of the challenges and limitations of current capabilities.

Jointly organized with Microsoft-CUHK Joint Laboratory for Human-Centric Computing and Interface Technologies.

Image Processing Using Quadratic Volterra Filters

by Professor Sanjit K. Mitra

Ming Hsieh Department of Electrical Engineering
University of Southern California, USA

Date: 4th December 2007, Tuesday


Abstract

Two-dimensional quadratic Volterra operators developed for edge enhancement are reviewed. A number of practical image processing applications of these filters are then considered and the proposed algorithms are outlined. These applications include image contrast enhancement, impulse noise removal, image zooming, and image half-toning. In each of these applications, it is shown that the processed images appear perceptually much better in quality than those obtained using many other well-known methods.

Sponsored Scholarly Events

IEEE International Conference on Nano/ Micro Engineered and Molecular Systems (IEEE-NEMS 2007)

January 16-19, 2007, Bangkok, Thailand

http://www.nanotec.or.th/ieee-nems2007/

W3C Voice Browser Working Group (VBWG) Internationalization of Speech Synthesis Markup Language (SSML) 1.1 Face-to-Face Meeting

July 11-13, 2006, The Chinese University of Hong Kong

Workshop on Optimization and Signal Processing (WOSP Hong Kong 2007)

December 19-21, 2007, The Chinese University of Hong Kong

http://www.se.cuhk.edu.hk/~zhang/WOSP2007/workshop.html

In this WOSP2007, 23 distinguished speakers were invited from all over the world to deliver technical talks in the emerging field of optimization and signal processing. They include:

Stephen Boyd,
Stanford University, USA

Raphael Cendrillon,
Hong Kong

Chong-yung Chi,
National Tsinghua University, Taiwan

Yonina Eldar,
Technion, Israel

Georgios Giannakis,
University of Minnesota, USA

Shunsuke Hayashi,
Kyoto University, Japan

Jian Li,
University of Florida, USA

Wing-Ken Ma,
CUHK, Hong Kong

Antonio De Maio,
University of Naples, Italy

Marc Moonen,
Katholieke Universiteit Leuven,
Belgium

Bjorn Ottersten,
KTH, Sweden

Asu Ozdaglar,
MIT, USA

Daniel Palomar,
HKUST, Hong Kong

Jong-Shi Pang,
University of Illinois, Urbana-
Champaign, USA

Liqun Qi,
Hong Kong Polytechnic University,
Hong Kong

Anthony So,
CUHK, Hong Kong

David Tse,
UC Berkeley, USA

Takashi Tsuchiya,
Institute of Statistical Mathematics,
Japan

Lieven Vandenberghe,
UCLA, USA

Dong Wei,
Huawei Technologies, China

Max Wong,
McMaster University, Canada

Yinyu Ye,
Stanford University, USA

Wotao Yin,
Rice University, USA

Award

Mr. Gary Chow Chun Tak, a student supported by the SHIAE funded project ("Systematic Synthesis of Bio-Nano-Informatics Chips"), has received a Croucher Scholarship to continue his PhD at Imperial College, UK. (April, 2007)

http://www.cse.cuhk.edu.hk/v5/about/hl/19.html

Contacts

For inquiries, please contact:

Professor P.C. Ching
Director, SHIAE

Tel: 2609-8275
Email: pcching@ee.cuhk.edu.hk

Professor Y.T. Zhang
Coordinator, Biomedical Engineering Track, SHIAE

Tel: 2609-8459
Email: ytzhang@ee.cuhk.edu.hk

Professor John C.S. Lui
Coordinator, Multimedia Technologies Track, SHIAE

Tel: 2609-8407
Email: cslui@cse.cuhk.edu.hk

Miss Natalie Tsang
Exective Assistant, SHIAE

Tel: 3163-4351
Email: ltsang@shiae.cuhk.edu.hk

William M.W. Mong Engineering Building
蒙民偉工程學大樓

Shun Hing Institution of Advanced Engineering (SHIAE)
The Chinese University of Hong Kong
Shatin, N.T., Hong Kong.

Office : Room 702, William M.W. Mong Engineering Building
Tel : (852) 3163 4351 Fax : (852) 3163 4354
Email : info@shiae.cuhk.edu.hk
Website : www.shiae.cuhk.edu.hk