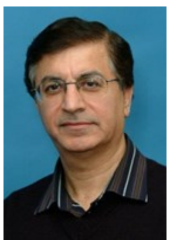
Issue 10

August 2015

APSIPA NEWSLETTER

Asia-Pacific Signal and Information Processing Association Newsletter

In This Issue



This is the 10th issue of APSIPA Newsletter which we crown it by the news of the very generous donation contributed by Professor Sadaoki Furui of US$250,000 to APSIPA. Prof Furui was the founding president of APSIPA (Asia-Pacific Signal and Information Processing Association), serving from 2009 to 2012. The donation will be used to promote and recognize high quality papers in the APSIPA Transactions on Signal and Information Processing by establishing a best paper award, named APSIPA Sadaoki Furui Prize Paper Award. Please refer to page 2 of this issue for more information. We also introduce interesting research articles to steer the readers to some stimulating research lines in quick shots. The articles included in this issue are:

1. ‘On adopting parametric array loudspeakers in active noise control’, by Chuang Shi, Yoshinobu Kajikawa, and Waleed H. Abdulla
2. ‘Phonated speech reconstruction: a short review’, by Hamid Sharifzadeh and Ian V. McLoughlin
3. ‘Joint modelling of linguistic and paralinguistic information – A new paradigm’, by Eliathamby Ambikairajah, Vidhyasaharan Sethu and Julien Epps
4. ‘SEAME: LDC release of Mandarin-English code-switching speech data’, by Ho Thi Nga, Chng Eng Siong, Haizhou Li
5. ‘Error signal measurement in active noise control systems’, by Darshana Sheth and Iman Ardekani

These are in addition to the regular columns such as the announcement to APSIPA conference, APSIPA Transactions updates, and others. We hope you find it useful and enjoying to read in your spare time. Please feel free to send us comments and suggestions to improve our newsletter to benefit as much as possible our APSIPA community.

Finally, to every beginning there is an end and this is the last issue to end up my term as Editor-in-Chief of APSIPA newsletter which I have been holding since the time I proposed it to our community. I would like to thank all the people who supported me to fulfill my commitments in this role and I will continue supporting the newsletter to become better and better.

Waleed Abdulla
APSIPA Newsletter Editor-in-Chief

Prof Sadaoki Furui Donates $25,000 ~~$250,000~~ to APSIPA

We are pleased to announce the donation of US$25,000 ~~US$250,000~~ to APSIPA from Prof Sadaoki Furui. Prof Furui was the founding president of APSIPA (Asia-Pacific Signal and Information Processing Association), serving from 2009 to 2012. The donation will be used to promote and recognize high quality papers in the APSIPA Transactions on Signal and Information Processing by establishing a best paper award, named **APSIPA Sadaoki Furui Prize Paper Award**.

Prof Furui has made significant contributions to the establishment of APSIPA. He first served as the steering committee chair from 2007 to 2009 bringing in many academic leaders in Asian-Pacific region with the vision to establish a signal and information association for this emerging and growing region. His effort has finally paid off with the establishment of APISPA at its first Annual Summit and Conference in October 2009. Prof. Furui was then elected and served as the founding President of APSIPA from 2009 to 2012, and has since continue to lead the development of APSIPA by serving as the Past President until 2014. Prof. Furui's vision and leadership brought APISAP from scratch to its current stage. His generosity and support of APSIPA will continue the legend that he has created for serving the APSIPA community.

Sadaoki Furui is currently the President of Toyota Technological Institute at Chicago, USA, and Professor Emeritus at Tokyo Institute of Technology. He is a world-recognized leader and has made significant contributions in a wide range of research on speech analysis, speech recognition, speaker recognition, speech synthesis, and multimodal human-computer interaction and has authored or coauthored over 900 published articles. He is a Fellow of the IEEE, the International Speech Communication Association (ISCA), the Institute of Electronics, Information and Communication Engineers of Japan (IEICE), and the Acoustical Society of America. He has served as President of the Acoustical Society of Japan (ASJ) and the ISCA. He has served as a member of the Board of Governors of the IEEE Signal Processing (SP) Society and Editor-in-Chief of both the Transaction of the IEICE and the Journal of Speech Communication. He has received the Yonezawa Prize, the Paper Award and the Achievement Award from the IEICE (1975, 88, 93, 2003, 2003, 2008), and the Sato

Prof. Furui the Golfer in Singapore

Paper Award from the ASJ (1985, 87). He has received the Senior Award and Society Award from the IEEE SP Society (1989, 2006), the Achievement Award from the Minister of Science and Technology and the Minister of Education, Japan (1989, 2006), the Purple Ribbon Medal from Japanese Emperor (2006), and the ISCA Medal (2009). In 1993 he served as an IEEE SPS Distinguished Lecturer.

Things you may not know about Professor Furui!

He is a good golfer. He plays with colleagues in japan and overseas. He is also a wonderful tennis player.

During APSIPA ASC 2013, APSIPA celebrated his career and he gave a 'retirement speech' to mark his retirement from Tokyo Institute of Technology shortly afterward he resumed a position as the President of Tokyo Technological Institute at Chicago.

As the founding president of APSIPA, he attended every APSIPA ASCs.

He is humble down to earth and never turn down any request for help.

Invitation to Participate in APSIPA ASC 2015

You are invited to participate in APSIPA ASC 2015 (www.apsipa2015.org), which will be held in Hong Kong from 16 to 19 December 2015. This conference is being organized by the Asia-Pacific Signal and Information Processing Association, and will serve as a forum to bring together experts in the areas of signal and information processing.

The conference will feature oral and poster presentations, talks by invited guests, as well as forums. The following are the keynote speeches for the conference.

Speech Title: **Whither Speech Recognition? - Deep Learning to Deep Thinking**
Prof. Sadaoki Furui, Toyota Technological Institute at Chicago
Speech Title: **Graph Signal Processing: Filterbanks, Sampling and Applications to Machine Learning and Video Coding**
Prof. Antonio Ortega, University of Southern California
Speech Title: **Can I Trust This Photo?**
Prof. Alex Kot, Nanyang Technological University

A Plenary Forum, titled "**The Future of Smart Life**", will also be organized for 18 December, 4:30pm -5:00pm. The speakers for this Forum include:
Yi Hao, President, TCL Multimedia, China
Kevin Jou, CTO and Corporate Vice President, MediaTek, Taiwan
Benoit Schillings, Vice President and Technical Fellow, Yahoo!, USA
Yasunori Mochizuki, Vice President, NEC Corporation, Japan
Qian Zhang (modulator), Professor, Hong Kong University of Science and Technology, Hong Kong

Eight tutorials will be offered on 16 December, which are free to all conference participants. Following are the details of the tutorials:

1. **Brain Computer Interface - overview of R&D and future outlook**

Sung Chan Jun, Gwangju Institute of Science and Technology, South Korea

2. **Assisted Listening for headphones and hearing aids: Signal Processing Techniques**
Woon-Seng Gan and *Jianjun He*, Nanyang Technological University, Singapore

3. **Introduction to Deep Learning and its applications in Computer Vision**
Wanli Ouyang and *Xiaogang Wang*, Chinese University of Hong Kong, Hong Kong

4. **Biomedical signal processing problems of human sleep monitoring**
Tomasz M. Rutkowski, University of Tsukuba, Japan

5. **Depth-based Video Processing Techniques for 3D Contents Generation**
Yo-Sung Ho, Gwangju Institute of Science and Technology (GIST), Korea

6. **Graph Signal Processing for Image Compression & Restoration**
Gene Cheung, National Institute of Informatics, Tokyo, Japan
Xianming Liu, Harbin Institute of Technology, China

7. **Spoofing and Anti-Spoofing: A Shared View of Speaker Verification, Speech Synthesis and Voice Conversion**
Zhizheng Wu, University of Edinburgh, UK
Tomi Kinnunen, University of Eastern Finland, Finland
Nicholas Evans, EURECOM, France
Junichi Yamagishi, University of Edinburgh, UK

8. **Wireless Human Monitoring**
Tomoaki Ohtsuki and *Jihoon Hong*, Keio University, Japan

Furthermore, more than 15 Invited Overview Sessions will be organized, in which renowned researchers will give overview talks on various research topics. In addition to regular oral and poster sessions, 22 special sessions on emerging research topics have been accepted for the conference. We expect around 400 delegates, from universities, research labs, companies and government agencies, to attend the conference.

We look forward to meeting you at the APSIPA conference in Hong Kong this December. You are also welcome to become an APSIPA member. To know more about APSIPA, please visit www.apsipa.org.

Niels Bohr

Born: October 07, 1885 in Copenhagen, Denmark,
Died: November 18, 1962
Niels Henrik David Bohr was a physicist who made foundational contributions to understanding atomic structure and quantum mechanics, for which he received the Nobel Prize in Physics in 1922. Bohr mentored and collaborated with many of the top physicists of the century at his institute in Copenhagen. He was part of a team of physicists working on the Manhattan Project. Bohr married Margrethe Nørlund in 1912, and one of their sons, Aage Bohr, grew up to be an important physicist who in 1975 also received the Nobel Prize. Bohr has been described as one of the most influential scientists of the 20th century.

Quoted:

“An expert is a person who has made all the mistakes that can be made in a very narrow field.”

On Adopting Parametric Array Loudspeakers in Active Noise Control

Chuang Shi[1], Yoshinobu Kajikawa[1], and Waleed H. Abdulla[2]
[1]Department of Electrical and Electronic Engineering, Kansai University
[2]Department of Electrical and Computer Engineering, University of Auckland

Abstract—The principal of parametric array loudspeaker (PAL) is based on transmitting an ultrasonic wave beyond the audible frequency range as a directional carrier to send audible sounds to a distant targeted location. PAL has been investigated to be used successfully in many applications. Ever increasing interests on the PAL are due to its sharp and controllable radiation pattern. In this article, we shade light over the use of PALs in active noise control (ANC), which is recognized as one of the promising applications to be widely proliferated.

I. PARAMETRIC ARRAY LOUDSPEAKER

Sound is one of the most important means by which we communicate with our external environments. The demand for customizing the sound field is growing in applications. The PAL provides an attractive new approach that is able to create a narrow sound beam from an ultrasonic emitter of a compact size. Acoustically speaking, the PAL is an application of the parametric acoustic array air. When two high pressure collimated ultrasonic beams are transmitted in the same direction, difference-frequency sound beam is generated in the transmission medium due to the nonlinear acoustic effect. Moreover, this difference-frequency beam is confined by the boundary of the ultrasonic beams. Thus, PAL is famous for its ability to generate a narrow sound beam from a thinner and smaller unit as compared to the conventional loudspeaker [1]. Furthermore, the steerable PAL is feasible to be realized by phased array techniques. The steerable PAL features an adjustable radiation pattern, which makes it suitable to be embraced in a variety of sound field control applications, including ANC.

II. FEASIBILITY AND VERSATILITY

Brooks, Zander and Hansen discussed the feasibility of using a PAL as the control source in ANC for the first time in 2005 [2]. They explained the motivation of considering a PAL. The conventional ANC system only minimized the sound pressure level (SPL) at the control point, which in turn increased the SPLs at other locations. This was known as the spillover problem of ANC. Using a PAL as the control source potentially solved the spillover problem. But the nonlinear nature of the PAL led to new problems. Brooks, Zander and Hansen listed three practical concerns, which were (1) the PAL was not good at producing low frequencies in terms of the SPL and harmonic distortion; (2) the precise control of amplitude and phase required for ANC was yet to be achieved by the PAL; (3) the ultrasonic exposure to the user might become a safety hazard.

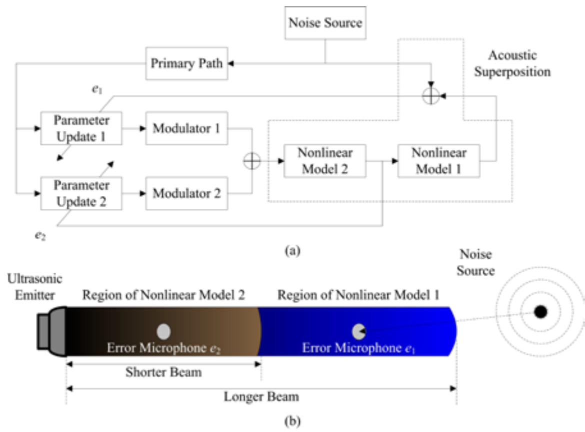


Fig. 1 Localized ANC using two PALs with different effective lengths (extracted and modified from [10]).

All the aforementioned concerns are basically due to the sound principle of the PAL. Since physical limitations are usually difficult to overcome, subsequent works on using PALs in ANC have been focused on the reduction of noisy frequencies in the midrange and upper midrange bands, such as from 500 Hz to 2.5 kHz. Unfortunately, the safety concern on the ultrasonic exposure has been marginalized.

Although there may be pessimistic views on using PALs in ANC, the related research is still undertaken. In 2006, Hansen's group proposed the integration of virtual sensing in ANC using a PAL [3] and carried out a superposition of the sound fields between a monopole source and a PAL in the numerical simulation [4]. An experiment validation was reported by Komatsuzaki and Iwata in 2008. However, the experimental results were not available in English until 2011 [5]. In Komatsuzaki and Iwata's setup, the primary noise source was a conventional loudspeaker producing a 1 kHz sine wave. The secondary control source was a PAL that was moved among three locations for investigation. Nevertheless the control was manual, the measured sound fields demonstrated the feasibility of solving the spillover problem by using a PAL in ANC.

Tanaka's group worked on the experimental validation too, and reported their results in local conferences in Japan. From 2008, they started to put effort into exploring unique features of the PAL that would improve the conventional ANC. Tachi and Tanaka [6] measured the noise reduction of an ANC system using the reflected sound of a PAL. In this special ANC system, the control point was invisible from the control source, which was infeasible for the conventional ANC. Then, a steerable PAL was designed to consist of 8 channels and steered the beam of a 1.5 kHz sound to 30° by controlling the phase of each channel. Experimental results of an ANC system using such a steerable PAL were published in 2010 [7]. In the following year, a curved-type PAL was built up for the purpose of global ANC by Tanaka's group [8]. The curved-type PAL had a focal point that was able to be freely configured to the exact location of a monopole noise source. Effective global noise reduction was observed in simulation and verified experimentally.

There are also some attempts to use more than one PAL in ANC at the same time. Nishiura's group considered the problem of the sound beam reflection when the PAL was used in a small room, and proposed to use another PAL to cancel the reflected sound beam of the first PAL [9]. Shi and Gan considered two coaxial PALs that had different effective lengths to form a length-limited sound spot that could further reduce the spillover of the conventional ANC [10]. The block diagram of using the length-limited sound spot in ANC is shown Fig. 1. It had been verified by a preliminary experiment in a duct to deal with a 1.5 kHz sine wave [11].

III. IMPLEMENTATION AND PROTOTYPING

Major advancement has been achieved in 2014 through adopting the digital signal processor (DSP) platforms in the controller design.

Yang's group worked out a feedforward controller for single-channel ANC [12]. This system was applied in an L-shape duct to reduce the fan noise. Using a PAL was found to give almost constant noise reduction over a long range in the duct. Panahi's group attempted to implement the leaky NLMS algorithm in Simulink and achieved good noise reduction performance for multi-tones [13].

Tanaka, Shi and Kajikawa investigated a factory noise problem, where the installation space for a conventional ANC system was limited. Since the PALs were able to be placed remotely from the control point, for example on the ceiling of the factory house, the ANC system using PALs became a

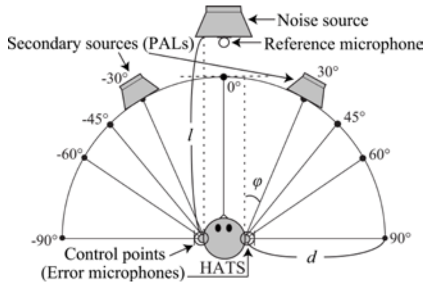


Fig. 2 Using two PALs to setup a case (1,2,2) multi-channel ANC (extracted from [14]).

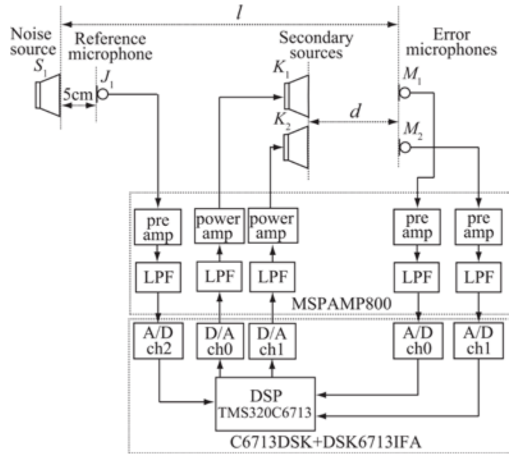
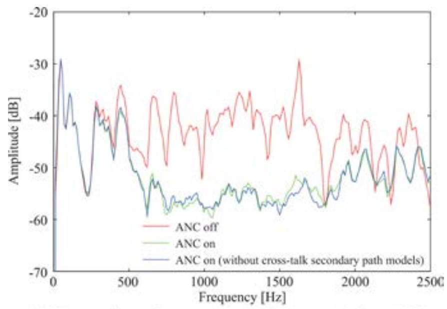
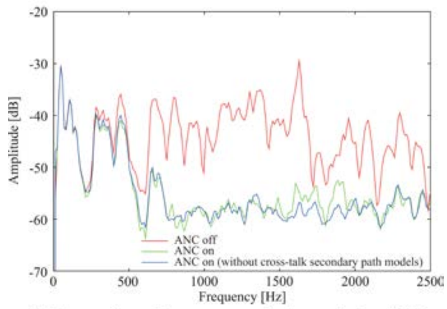


Fig. 3 DSP implementation of the case (1,2,2) multi-channel ANC using PALs (extracted from [15]).



(a) Comparison of error spectra before and after ANC (left error microphone location)



(b) Comparison of error spectra before and after ANC (right error microphone location)

Fig. 4 Noise reduction performance of the case (1,2,2) multi-channel ANC using PALs (extracted from [15]).

favorable solution. Moreover, two control points were necessary for each person. Therefore, the case (1,2,2) multiple-channel ANC system was chosen, of which the setup and DSP implementation are shown in Figs. 2 and 3 [14, 15]. There is a computational advantage of the case (1,2,2) multi-channel ANC system using PALs. For the PALs are highly directional sound sources, the cross-talk secondary paths are negligible. Experimental results show that noise reduction performance is almost not affected by cross-talk secondary path models (see Fig. 4) [15].

IV. FUTURE RESEARCH SUGGESTION

Using PALs in ANC is an interesting research topic and potentially becomes a special solution to a lot of real word noise problems. In the past studies, more and more unique features of using PALs have been discovered. Therefore, an emerging question is how those features can be derived from the nonlinear acoustic theory of the PAL [16]. So far, the integration of the PAL and ANC is superficial to the extent that the PAL is treated as a directional loudspeaker. In fact, the PAL is a complicated nonlinear device and has its own limitations to be solved. The future works on using PALs in ANC should contain three aspects. Firstly, the in-depth integration of the PAL and ANC is expected to find the optimized application instead of the optimized solution, accounting for the existing limitations of PALs. Secondly, psychoacoustic approaches are beneficial to overcome the physical constraints [17]. The psychoacoustical ANC has already become a popular research problem regardless of using PALs. Lastly, we have hope on the advancement of the PAL to possess a linear response to add synergy to its deployment in ANC as well as other sound field control applications.

REFERENCE

[1] C. Shi, Y. Kajikawa, and W. S. Gan, "An overview of directivity control methods of the parametric array loudspeaker," *APSIPA Transactions on Signal and Information Processing*, vol. E20, no. 3, 2014.

[2] L. A. Brooks, A. C. Zander, and C. H. Hansen, "Investigation into the feasibility of using a parametric array control source in an active noise control system," *Proceedings of Acoustics 2005*, Busselton, Australia, 2005.

[3] M. R. F. Kidner, C. Petersen, A. C. Zander, and C. H. Hansen, "Feasibility study of localised active noise control using an audio spotlight and virtual sensors," *Proceedings of Acoustics 2006*, Christchurch, New Zealand, 2006.

[4] M. R. F. Kidner, A. C. Zander, and C. H. Hansen, "Active control of sound using a parametric array," *Proceedings of Active 2006*, Adelaide, Australia, 2006.

[5] T. Komatsuzaki and Y. Iwata, "Active noise control using high-directional parametric loudspeaker," *Journal of Environment and Engineering*, vol. 6, no. 1, pp. 140-149, 2011.

[6] R. Tachi and N. Tanaka, "Active noise control using acoustic wave reflection of parametric speaker," *Proceedings of APVC 2009*, Christchurch, New Zealand, 2009.

[7] N. Tanaka and M. Tanaka, "Active noise control using a steerable parametric array loudspeaker," *Journal of the Acoustical Society of America*, vol. 127, no. 6, pp. 3526-3537, 2010.

[8] N. Tanaka and M. Tanaka, "Mathematically trivial control of sound using a parametric beam focusing source," *Journal of the Acoustical Society of America*, vol. 129, no. 1, pp. 165-172, 2011.

[9] S. Kurimoto, M. Morise, and T. Nishiura, "The suppression for undesired reflection towards audio spot," *Proceedings of ICA 2010*, Sydney, Australia, 2010.

[10] C. Shi and W. S. Gan, "Using the length-limited parametric source in active noise control applications," *Proceedings of the 20th ICSV*, Bangkok, Thailand, 2013.

[11] B. Lam, W. S. Gan, and C. Shi, "Feasibility of a length-limited parametric source for active noise control applications," *Proceedings of the 21st ICSV*, Beijing, China, 2014.

[12] S. Wu, M. Wu, and J. Yang, "An investigation of parametric array loudspeakers for active noise control in a L-shaped duct," *Proceedings of the 21st ICSV*, Beijing, China, 2014.

[13] A. Ganguly, S. H. K. Vemuri, and I. Panahi, "Real-time remote cancellation of multi-tones in an extended acoustic cavity using directional ultrasonic loudspeaker," *Proceedings of IECON 2014*, Dallas, Texas, 2014.

[14] K. Tanaka, C. Shi, and Y. Kajikawa, "Study on active noise control systems using parametric array loudspeakers," *Proceedings of Forum Acusticum 2014*, Krakow, Poland, 2014.

[15] K. Tanaka, C. Shi, and Y. Kajikawa, "Multi-channel active noise control using parametric array loudspeakers," *Proceedings of APSIPA 2014*, Siem Reap, Cambodia, 2014.

[16] W. K. Tseng, "Quiet zone design in diffuse fields using ultrasonic transducers," *Journal of Applied Mathematics and Physics*, vol. 3, no. 2, pp. 247-253, 2015.

[17] T. Tamesue and T. Saeki, "Sound masking for achieving speech privacy with parametric acoustic array speaker," *Proceedings of SCIS & ISIS 2014*, Kitakyushu, Japan, 2014.

Phonated Speech Reconstruction: A Short Review

Hamid R. Sharifzadeh*, Ian V. McLoughlin†
* Department of Computing, Unitec Institute of Technology, Auckland, New Zealand
E-mail: hsharifzadeh@unitec.ac.nz
† School of Computer Science, The University of Kent, Kent, United Kingdom
E-mail: i.v.mcloughlin@kent.ac.uk

***Abstract*—Whispered speech is useful for quiet and private communications in daily life but becomes the primary communicative mechanism for many people experiencing voice box difficulties, such as post-laryngectomised patients. Computational speech reconstruction algorithms not only can help voice-impaired individuals to speak with natural speech (which is not possible with current rehabilitation methods), but also can be used by unimpaired speakers for communicating sensitive or private information.**

In this short review, we will provide a brief summary on some of whispers to phonated speech reconstruction algorithms. Based on our recent paper published in ACM Transactions on Accessible Computing, such reconstruction methods are discussed and the efficiency of the most recent method in terms of speech reproduction quality and computational complexity is briefly outlined.

I. Introduction

The human voice is the most magnificent instrument for communication, capable of expressing deep emotions, conveying oral history through generations, or of starting a war. However, those who suffer from aphonia (no voice) and dysphonia (voice disorders) are unable to make use of this critical form of communication. They are typically unable to project anything more than hoarse whispers [1], [2], [3]. Furthermore, as secondary means of communications, whispers play a significant role in everyday speech [4] for communicating sensitive or private information, or when speaking in locations such as libraries, during lectures and meetings, in which normal speech may be deprecated.

Even though most speakers will whisper at times, and some speakers can only whisper, the majority of today's computational speech technology systems assume or require phonated speech. Current ASR systems, speech communications devices, voice authentication devices and speech input technologies tend to be either incapable of handling whispers, or capable of only degraded operation with whispers [5], despite whispers being a natural and integral part of face-to-face communications between humans. Once whispers can reliably be reconstructed into phonated speech, it is possible that speech-only devices and systems could then work with whispers without requiring modification. On the other hand, such reconstruction approaches would also be useful for medically-related whispers in which current rehabilitative procedures (or devices) such as esophageal speech [6], transoesophageal puncture (TEP) [7], and electrolarynx [8] suffer from various drawbacks ranging from difficulty of usage and risk of infection to generation of monotonous or robotised speech [9], [10], [11].

Several computational methods for converting whispers to normal speech have been introduced recently [12], [13], [14], [15], [16], [17], [18]. These speech reconstruction algorithms can be classified into two major groups of those requiring training and those which do not. Utilising machine learning algorithms are the basis of training-based methods (whispers are mapped to corresponding normal speech) while non-training methods rely upon whisper enhancement and pitch regeneration. In this short paper, we overview the state of the art in whisper-to-speech reconstruction by summarising our recent publication [15].

II. Computational Speech Reconstruction Methods

The pioneering whisper-to-speech conversion approach is the mixed excitation linear prediction (MELP) based system of Morris et al. [13]. The method requires parallel same-speaker training (i.e., both normal and whispered recordings) for a jump Markov linear system which then estimates pitch and voicing parameters. The technique reportedly works well, however its main weaknesses are that it cannot be used for speakers whose original voice has already been lost, and that the technique is not well suited for real-time operation.

In order to overcome these limitations, a code-excited linear predictor (CELP) based alternative was subsequently proposed [12]. This derives pitch excitation from a selection of fixed pitch models instead of training individual models for each speaker. Being similar to a CELP decoder (i.e., CELP without the codebook search loop), it was potentially suitable for real-time operation in terms of complexity.

Both the MELP and CELP methods were shown to work well for phonemes, diphthongs and single-words, but neither claimed to work or were fully evaluated for continuous whisper-to-speech reconstruction.

Statistical voice conversion (SVC) approaches for reconstruction have emerged more recently. Most notable are the systems developed by Toda et al. [16], [18] which make use of Gaussian-mixture models (GMM) to independently model pitch contours and spectral parameters from parallel whisper/speech training data. In fact, Toda et al. began by converting non-audible murmur (NAM) signals into realistic sounding speech, and then extended the system to convert whispers.

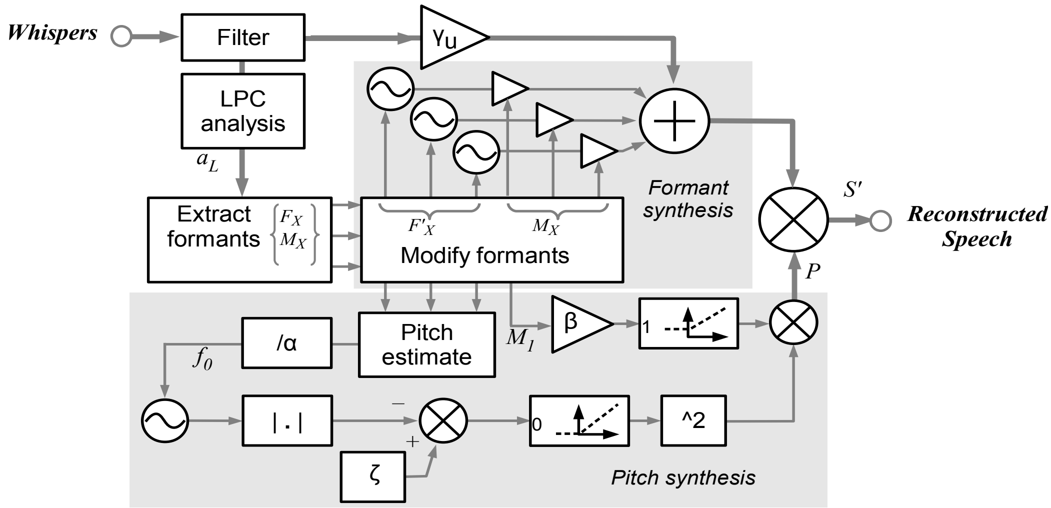


Fig. 1. Block diagram of sine wave speech based reconstruction mechanism with enhanced pitch modulation using inter-formant harmonic relationship.

These methods are capable of transforming whisper acoustic features into those more resembling natural speech after being suitably trained with parallel utterance data (i.e., speech and whisper recordings of the same speech by the same speaker). In general, three GMMs are used: one converts source spectral features into target spectral features, another converts the same source spectral features into a pitch (or f_0) feature. The final GMM generates target aperiodic components, which are useful for preserving naturalness. Reconstruction relies upon STRAIGHT [19], making use of estimated f_0, spectral and aperiodic components. Although the quality of reconstructed speech is high, these methods suffer from over-smoothing which tends to remove or muffle detailed characteristics in the resulting spectra. At times, there may also be an unnatural prosody due to the difficulty of estimating f_0 from whisper spectral features [14].

However, among all current methods, SVC systems probably yield the highest quality of reconstructed speech from whisper input. Unfortunately, they suffer from two major disadvantages. The first is that, similar to the MELP method of Morris et al. [13], significant amounts of clean speech and corresponding whisper utterances are required *a priori* to train the system, and then the resulting trained models are specific to one speaker only. The second disadvantage is that the entire process of reconstruction involves quite significant computational overheads given that multiple GMMs (or restricted Boltzmann machines [14]) are required, synthesis requires additional software packages, and the entire process requires highly overlapped analysis frames (typically 20 ms in size, advancing at 5 ms each iteration).

Considering methods that do not require user-specific *a priori* information, a new approach based on mixing modulated sine wave with scaled original whisper input was introduced recently by McLoughlin et al.[20], [15]. This method does not require training, and does not require the availably of parallel speech/whisper input data. The method, shown diagrammatically in Fig. 1, defines a harmonic relationship between pitch and formants to synthesise a pseudo-f_0. The enhanced approach [15] makes use of higher formant information (when present), specifically the inter-formant harmonic relationship, to directly modulate sine wave speech while the original system [20] constrains f_0 to be a fixed integer sub-multiple of only first formant (F1). Importantly, there is no voiced/unvoiced (V/UV) or formant switching employed in the system, since determining V/UV status from whispers proves to be a difficult and error-prone task.

The proposed system was configured to track and reconstruct up to $N_s = 4$ formant candidates per analysis window. To set up the system for speech regeneration, the 128 sample analysis windows were highly overlapped by 87.5% and formant extraction LPC analysis order was 8, and sample rate set to 8 kHz. As an example, waveforms and spectrograms are plotted for spoken, whisper and reconstructed speech for the sentence "*Should giraffes be kept in small zoos?*" in Fig. 2. The reconstructed spectrogram shows wider and more prominent formant bands (for example, compare the /a/ between 2.1 and 2.4 seconds in the bottom two spectrograms with the corresponding phoneme from the top spectrogram, located between 1.8 and 2.1 seconds). The reconstructed speech has a better lower frequency energy distribution, although it still lacks much of the pitch modulation energy that is present in the original speech.

III. Conclusion

This short paper briefly reviewed some of whisper-to-speech reconstruction algorithms and highlighted a recent approach based on 'sinewave speech' technique. This approach was enhanced with novel pitch regeneration mechanism aiming to convert whispers to natural-sounding voiced speech. This parametric reconstructor does not require training or access to any *a priori* information. Further details on this approach including objective and subjective evaluations and comparison with output of other reconstruction algorithms can be found in [15].

Despite improvements in the quality of reconstructed speech in recent years, more research is still required in this field. People using these voice reconstruction techniques will have

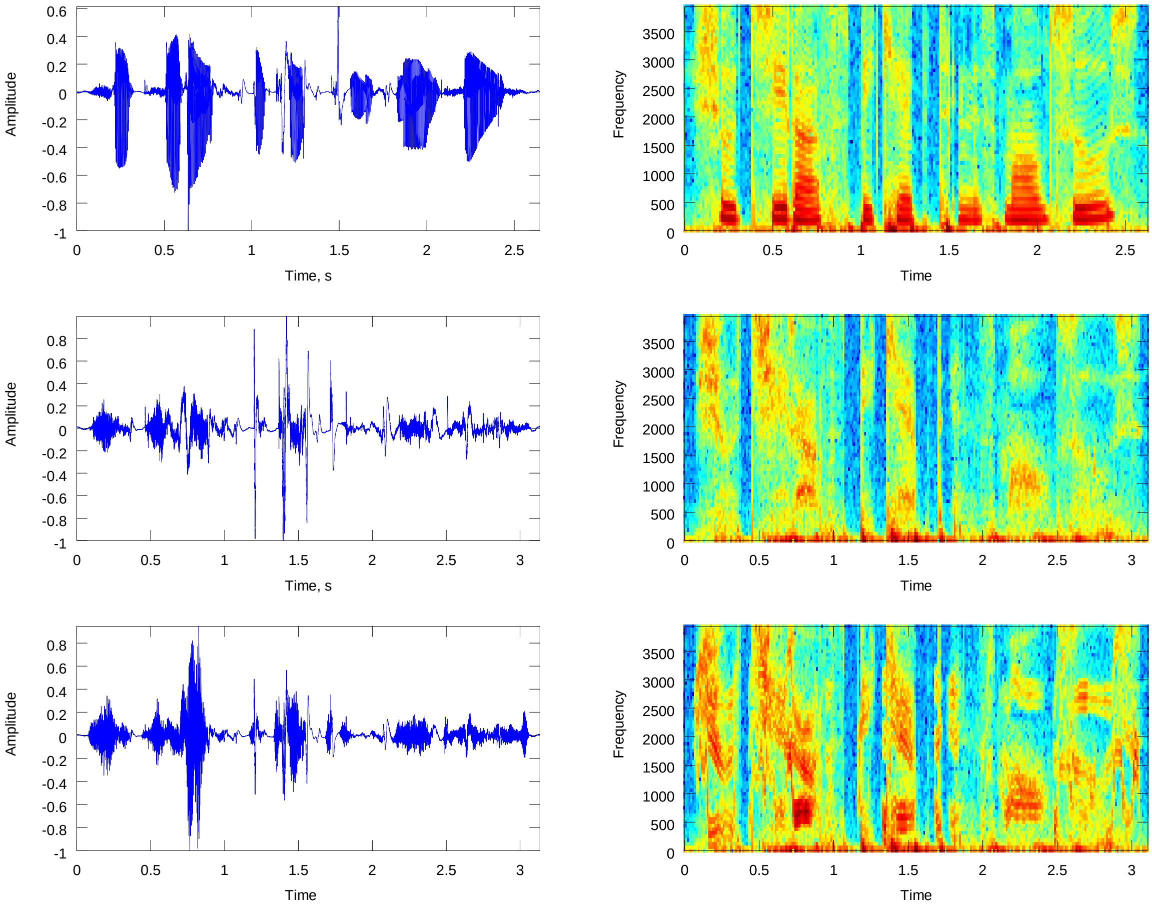


Fig. 2. Waveform and spectrogram plots of the sentence "*Should giraffes be kept in small zoos?*" showing (top) spoken, (middle) whispered and (bottom) reconstructed speech. All are amplitude normalised prior to plotting.

the benefit of a reconstructed voice, but not yet one which has met the desired target of being natural-sounding.

REFERENCES

[1] R. Pietruch, M. Michalska, W. Konopka, and A. Grzanka, "Methods for formant extraction in speech of patients after total laryngectomy," *Biomedical Signal Processing and Control*, vol. 1, pp. 107–112, 2006.

[2] H. R. Sharifzadeh, I. V. McLoughlin, and M. J. Russell, "A comprehensive vowel space for whispered speech," *Journal of Voice*, vol. 26, no. 2, pp. e49 – e56, 2012.

[3] N. P. Solomon, G. N. McCall, M. W. Trosset, and W. C. Gray, "Laryngeal configuration and constriction during two types of whispering," *Journal of Speech and Hearing Research*, vol. 32, pp. 161–174, 1989.

[4] V. C. Tartter, "Whats in a whisper?" *Journal of the Acoustical Society of America*, vol. 86, pp. 1678–1683, 1989.

[5] H. Beigi, *Speaker Recognition: Advancements and Challenges*. Intech, 2012.

[6] M. Azzarello, B. A. Breteque, R. Garrel, and A. Giovanni, "Determination of oesophageal speech intelligibility using an articulation assessment," *Revue de laryngologie, otologie, rhinologie*, vol. 126, pp. 327–334, 2005.

[7] V. Callanan, P. Gurr, D. Baldwin, M. White-Thompson, J. Beckinsale, and J. Bennet, "Provox valve use for post-laryngectomy voice rehabilitation," *Journal of Laryngology and Otology*, vol. 109, pp. 1068–1071, 1995.

[8] J. H. Brandenburg, "Vocal rehabilitation after laryngectomy," *Archives of Otolaryngology*, vol. 106, pp. 688–691, 1980.

[9] G. Culton and J. Gerwin, "Current trends in laryngectomy rehabilitation: A survey of speech language pathologists," *Otolaryngology - Head and Neck Surgery*, vol. 115, pp. 458–463, 1998.

[10] H. R. Sharifzadeh, I. V. McLoughlin, and F. Ahmadi, *Lecture Notes in Electrical Engineering*. Springer, 2010, ch. Speech rehabilitation methods for laryngectomised patients, pp. 597 – 607.

[11] H. Liu, Q. Zhao, M. Wan, and S. Wang, "Enhancement of electrolarynx speech based on auditory masking," *IEEE Transactions on Biomedical Engineering*, vol. 53, pp. 865–874, 2006.

[12] H. R. Sharifzadeh, I. V. McLoughlin, and F. Ahmadi, "Reconstruction of normal sounding speech for laryngectomy patients through a modified celp codec," *IEEE Transactions on Biomedical Engineering*, vol. 57, pp. 2448–2458, 2010.

[13] R. W. Morris and M. A. Clements, "Reconstruction of speech from whispers," *Medical Engineering and & Physics*, vol. 24, pp. 515 – 520, 2002.

[14] J. Li, I. V. McLoughlin, L. Dai, and Z. Ling, "Whisper-to-speech conversion using restricted boltzmann machine arrays," *Electronics Letters*, vol. 50, no. 24, pp. 1781 – 1782, 2014.

[15] I. V. McLoughlin, H. R. Sharifzadeh, S. Tan, J. Li, and Y. Song, "Reconstruction of phonated speech from whispers using formant-derived plausible pitch modulation," *ACM Transactions on Accessible Computing*, vol. 6, no. 4, pp. 12:1–12:21, 2015.

[16] T. Toda, M. Nakagiri, and K. Shikano, "Statistical voice conversion techniques for body-conducted unvoiced speech enhancement," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, no. 9, pp. 2505 – 2517, 2012.

[17] H. R. Sharifzadeh, I. V. McLoughlin, and F. Ahamdi, "Voiced speech from whispers for post-laryngectomised patients," *IAENG International Journal of Computer Science*, vol. 36, no. 4, pp. 367 – 377, 2009.

[18] T. Toda and K. Shikano, "Nam-to-speech conversion with gaussian mixture models," in *INTERSPEECH*, 2005.

[19] H. Kawahara, M. Morise, T. Takahashi, R. Nisimura, T. Irino, and H. Banno, "Tandem-straight: A temporally stable power spectral representation for periodic signals and applications to interference-free spectrum, f0, and aperiodicity estimation," in *ICASSP*, 2008.

[20] I. V. McLoughlin, J. Li, and Y. Song, "Reconstruction of continuous voiced speech from whispers," in *INTERSPEECH*, 2013.

Joint modelling of linguistic and paralinguistic information – A new paradigm

Eliathamby Ambikairajah, Vidhyasaharan Sethu and Julien Epps

School of Electrical Engineering and Telecommunications

UNSW Australia

The primary objective of speech is to convey linguistic information via one of the spoken languages; however the speaker's physiological, emotional, mental and cognitive states influence the speech characteristics and these states are collectively referred to as paralinguistic information. Humans are able to convey linguistic and interpret paralinguistic information in speech with very little effort during a normal conversation. For example, we can infer from speech what is being said, what language is being spoken, who is speaking, gender of the speaker, approximate age of the speaker, emotional state, stress level, cognitive load level, depression level etc (see Figure 1).

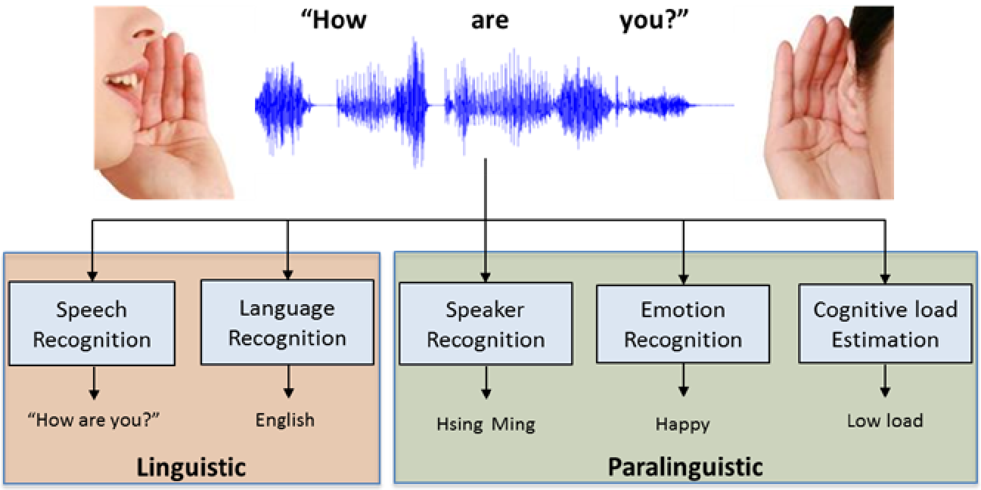


Figure 1: Speech based information recognition systems

Intuitively, one imagines that the detailed linguistic and paralinguistic content of speech is jointly encoded and decoded, however when dealing with specific speech-based recognition applications on an individual basis, it has usually been convenient to ignore the interdependence and implement an individual system for the application. However, it is important to recognise that speech recognition, speaker recognition and emotion recognition systems are all ultimately part of a single user interface, and from a user's perspective in principle it would be better if the system were more aware of all linguistic and paralinguistic information rather than ignoring some of the content. New insights and new approaches are needed to combine linguistic and paralingustic information in implementing any speech based system.

Present day speech and speaker recognition technologies are able to cope with these subtle paralinguistic variations only in a rather limited manner. For instance, automatic speech recognition systems are not capable of modelling paralinguistic information and consequently their performance degrades unless restrictions are placed on vocabulary, speaker, speaking style, etc. Most state of the art systems normalise these paralinguistic variables prior to pattern matching to improve robustness. However the scope of such normalisation for removing variability due to paralinguistic information in acoustic features is limited, and if there is significant paralinguistic variability then system performance typically degrades as a result. Nevertheless contemporary speech and speaker recognition systems employ a range of normalisation, warping and modelling techniques that compensate for variability in channel effects, emotional state of the speaker, stress level, etc .

Future speech based systems should jointly model linguistic and information. That is, integrate the information obtained by various traditionally independent systems such as speech recognisers, speaker recognisers, emotion recognisers and cognitive load estimators under a common framework. A conceptual model of a joint linguistic and paralinguistic classification system is shown in Figure 2. This general approach is motivated by the success of the individual classification systems that are prevalent in current literature. It aims to incorporate the different information obtained by individual systems that have been optimised for their respective applications, and to make use of them to aid in the joint modelling.

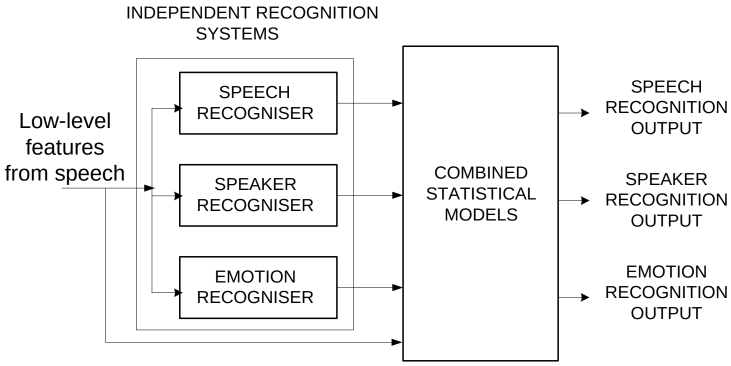


Figure 2: Joint linguistic and paralinguistic modelling approach

The general idea behind this joint modelling is that the individual recognition systems can be leveraged to generate reasonable N-best estimates of the target values, e.g., N-best word estimates from the speech recogniser, N-best speaker identity estimates, N-best emotional class estimates, etc. In addition, the individual recognition system can also generate the likelihood values (or equivalently scores) corresponding to these N-best estimates. Suitable statistical models of the original feature space and inference methods that leverage these N-best estimates and likelihoods can then be used to jointly and more accurately estimate both linguistic and paralinguistic information. Joint modelling of linguistic and paralinguistic information as outlined here is a new paradigm for speech based recognition systems and opens up multiple avenues of research. A number of challenges exist in realising this approach, including the lack of any single database that has all relevant linguistic and paralinguistic labels, the lack of any single database that is sufficiently large to robustly estimate the large number of complex models required, and the identification of suitable machine learning techniques that are able to robustly model this multi-dimensional information with sufficient complexity given the paucity of data.

SEAME: LDC release of Mandarin-English code-switching speech data

Ho Thi Nga[1], Chng Eng Siong[1], Haizhou Li[1,2]
[1]Nanyang Technological University, Singapore
[2]Institute for Infocomm Research, Singapore

SEAME (South-East Asia Mandarin-English) is the first publicly released Mandarin-English code-switching speech database for automatic speech recognition (ASR) research in LDC. The database was jointly developed by the Speech Groups in Temasek Laboratories@Nanyang Technological University, Singapore and Universiti Sains Malaysia, Malaysia. It was released in April 2015 through Linguistic Data Consortium or LDC (corpus id: LDC2015S04) [5]. The SEAME corpus includes 192 hours of Mandarin-English code-switching and mono-lingual utterances under conversation and interview settings.

What is code-switching speech?

Code-switching refers to the practice of shifting between languages or language varieties during conversation. Such behaviour in a conversation has become more common in bilingual societies, e.g., in Singapore and Malaysia, where the population consists of multiple races who can speak their respective native language well along with the common working language English. It is hence quite common for the people to mix English words or phrases with those of other languages [1].

Developing a code-switching ASR system is considerably more difficult than a mono-lingual system. This is because the code-switching system will need to accommodate various languages pronunciation and language model simultaneously. Furthermore, there are only very few code-switching corpora available publicly to develop such systems. In addition, available corpuses are usually small in scale [1].

What can we do with the SEAME corpus?

The SEAME corpus is designed to support many aspects of code-switching studies related to ASR:
- Pronunciation modelling
- Language modelling
- Acoustic modelling
- Automatic speech recognition
- Language identification [2]
- Language turn detection [3][4]

SEAME Database release

The SEAME (LDC2015S04) corpus was first released by LDC on 15 April 2015. It consists of 192 hours of speech and 63 hours of word level manual transcription. Most of the 63 hours of transcribed speech are code-switching utterances. The un-transcribed speech data are mostly mono-lingual segments and its transcription will be released in Q3 of 2015.

The corpus was recorded with two speaking styles, conversational and interview, 92 hours for conversational style and about 100 hours for interview style. In a conversation, two speakers are involved. They conduct a dialogue freely with each other when their speech is recorded through closed talk microphones. In an interview, an interviewer asks general questions about sports, daily lifestyles, etc., and only the interviewee’s response is recorded through close talk microphone. The corpus was recorded in two sites (Singapore and Malaysia) with the same recording setup during

2009 to 2010, with a close-talk microphone in a quiet environment. Each recording session is about one hour long for either interview or conversation. There are 156 distinct speakers in total, balanced in gender and their ages are between 19 and 33.

Where to get the database?

Visit LDC website: https://catalog.ldc.upenn.edu/LDC2015S04.

Second release of SEAME

An update of SEAME corpus will be released by LDC in Q3, 2015 with the same corpus ID. In this update, the mono-lingual portion of the SEAME recording, which was not transcribed in the first release, will be made available.

References

[1] Dau-Cheng Lyu, Tien Ping Tan, Engsiong Chng, Haizhou Li: SEAME: a Mandarin-English code-switching speech corpus in south-east asia. INTERSPEECH 2010: 1986-1989

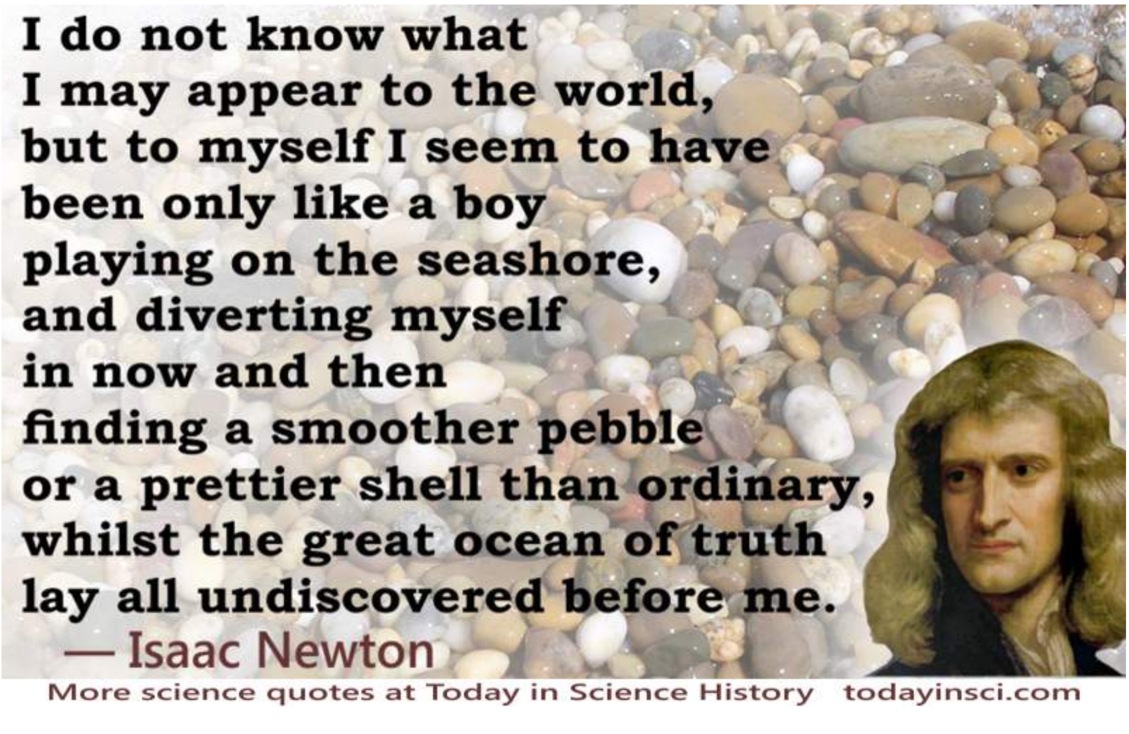
[2] Ngoc Thang Vu, Dau-Cheng Lyu, Jochen Weiner, Dominic Telaar, Tim Schlippe, Fabian Blaicher, Engsiong Chng, Tanja Schultz, Haizhou Li: A first speech recognition system for Mandarin-English code-switch conversational speech. ICASSP 2012: 4889-4892

[3] Dau-Cheng Lyu, Eng-siong Chng, Haizhou Li: Language diarization for code-switch conversational speech with pronunciation dictionary adaptation. ChinaSIP 2013

[4] Dau-Cheng Lyu, Tien-Ping Tan, Eng-Siong Chng, and Haizhou Li: An Analysis of a Mandarin-English Code-switching Speech Corpus: SEAME. Oriental COCOSDA 2010.

[5] https://catalog.ldc.upenn.edu/LDC2015S04

[6] Dau-Cheng Lyu, Tien-Ping Tan, Eng-Siong Chng, and Haizhou Li : Mandarin–English code-switching speech corpus in South-East Asia: SEAME. Language Resources and Evaluation *(To appear)*



Error Signal Measurement in Active Noise Control Systems

Darshana Sheth and Iman Ardekani
Department of Computing, Unitec Institute of Technology, Auckland, New Zealand

Abstract: Active noise control (ANC) creates a set of silent points at the location of error microphones. As a byproduct, a zone of quiet is created surrounding the silent points. Unfortunately, the extension of the created quite zone is small. Moreover, a large part of the quiet zone is occupied by the error microphones. Recently, a number of researchers have independently proposed to displace error microphones and move them outside the desired quite zones. This idea involves in the modification of ANC adaptive algorithms. This paper looks at available techniques that aims at the displacement of error microphones in ANC systems and conducts a comparison discussion about them.

I. Introduction

ANC utilizes an electroacoustic framework to cancel out the undesirable sound (noise) based on the superposition principle. In an ideal ANC system, a control system drives a loudspeaker to create a sound field that is equal in magnitude but opposite in phase with the original noise field. The superposition of the two fields results in silence. In practice, this idea can be only realized for a set a discrete points in the field. ANC has been focused in the control of low frequency noise, where the passive noise control systems are inefficient due to this fact that the noise wavelength is comparable with the dimensions of the acoustic barriers. ANC systems use adaptive filters to derive the cancelling loudspeaker. The most common adaptive filters used for ANC consists of a finite impulse response (FIR) filter with an LMS-type algorithm. In view point of control theory, ANC systems are classified in to two categories: feedforward and feedback. Feedforward ANC enjoys the knowledge of the noise field obtained through measurement in a location close to the noise source. Feedback ANC must estimate the noise field by using an internal model.

As shown in Figure 1, a typical feedforward ANC system has a single reference microphone, a single cancelling loudspeaker, and a single error microphone. The reference microphone picks up a reference signal *x(n)*. *x(n)* is filtered by the ANC adaptive filter to produce the anti-noise signal *y(n)* which is fed to the loudspeaker. The error microphone picks up the error signal *e(n)* which is the combination of the noise and anti-noise [1]. Feedforward ANC is usually more reliable than feedback ANC.

Error microphones play a significant role in ANC systems. The reference microphones can be removed from the system in feedback ANC. However, both feedforward and feedback ANC rely on the error signal

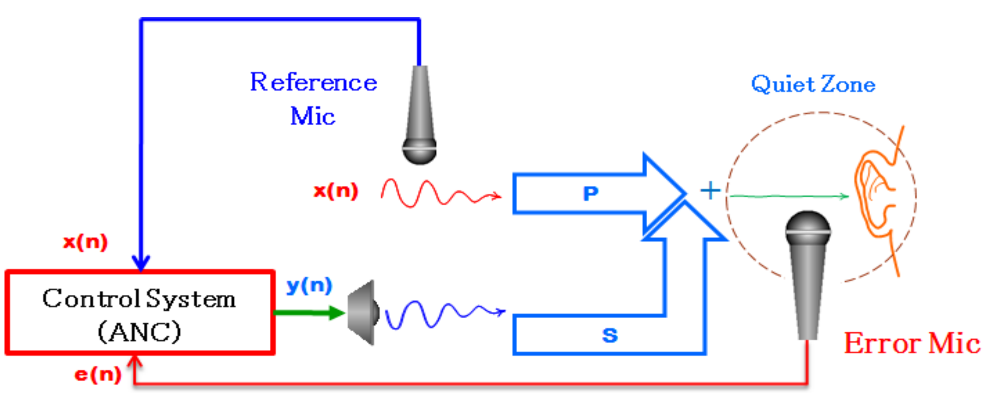


Figure 1. Single channel feedforward ANC

e(n) picked up by the error microphone [2]. ANC adaptive algorithms computes an optimal value for *y(n)* signal through the minimization of *e(n)* [3].
Filter-x Least Mean Square (FxLMS) is the most common ANC adaptive algorithm. Different variants of the FxLMS, for example, Filtered-x Normalized LMS (FxNLMS), Leaky FxLMS, Modified FxLMS (MFxLMS) were proposed to enhance the performance of the original algorithm. One common shortcoming of FxLMS and its variants is the slow convergence rate. To overcome this shortcoming, more complicated algorithms such as Filtered-x Recursive Least Square (FxRLS) or Filtered-x Affine Projection (FxAP) can be used. These algorithms have faster convergence rate at the cost of computational complexity.

II. Different Approaches to Remote ANC

In many applications, error microphones can't be placed close inside the quite zone (e.g., close to the audience's ears). To form a quiet zone far from error microphone, different techniques have been introduced. In the following those techniques are discussed.

A. Moving microphone

Another solution might be based on using a moving error microphone. When the ANC system is tracking moving error microphones, the adaptive algorithm has to be parameterized in a different way. The adaptation must be fast enough to track the movements of the error microphone [4]. The proposed ANC system with moving error microphone, is able to create 3-dimensional zones of quiet in an enclosure. As shown

in Figure 2, the quiet zone is created around the error microphone *Me*. The rotational speed can range from about 0.2 to 1.3 rotations per second. Based on the results reported in [4], the mean attenuation in the error microphone calcu
algorithms were proposed in [4] to realize the ANC system with moving error microphones. However, these algorithms have nearly similar performance, as shown in Table 1.

B. Remote ANC

The idea of remote ANC is proposed in [3] and [8]. The reference signal is independent of the control system. Therefore, the reference signal of the remote ANC system is identical to that of the original FxLMS based ANC system. The error signal in the remote ANC

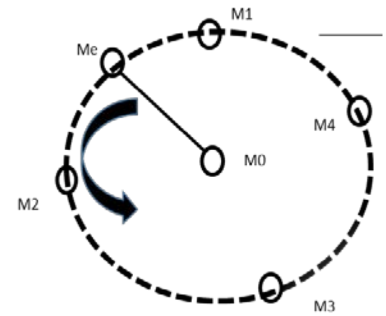


Figure 2. ANC with moving error microphone [4]

Table 1. noise attenuation for moving microphone

Algorithm	V=0.5m/s	V=2.1m/s	V=3.7m/s
FX-LMS	28db	19db	14db
NFX-LMS	28db	20db	15db
MFX-LMS	29db	20db	16db

Table 2. noise attenuation for remote ANC

Algorithm	r=0cm	r=2cm	r=4cm
R FX-LMS	22db	20db	19db

system is different with the error signal in the original FxLMS based ANC system [8]. A mechanism for the compensation of this this difference was proposed in [3] and [8], resulting in a new ANC algorithm, called Remote FxLMS (R FxLMS). Table 2 shows the simulation results obtained in different experiments. In each experiment a particular distance between the actual microphone and the location of interest is used. As seen, by moving the microphone further, the performance of remote ANC algorithm is degraded.

C. Virtual microphone

As the zone of quiet produced at the physical error microphone is restricted in size for ANC, virtual acoustic sensors were created to move the zone of quiet to a desired area that is remote from physical error microphone. Utilizing the physical error signal, a virtual detecting algorithm is utilized to scale the weight at an improved virtual area. These moving virtual sensing algorithms evaluate the error signals at various virtual areas that travel through the sound field. Various moving virtual sensing algorithms have been produced including the spatially fixed virtual sensing algorithm, the remote moving error microphone technique [7], the adaptive LMS moving virtual error microphone technique [9] and the Kalman filtering moving virtual sensing technique [10].

ANC systems using this technique can be classified in to two main categories: model-based and non-model-based. The former requires a model of the acoustic plant to process data picked up from microphones. An off-line acoustic modeling technique is necessary to reach the acoustic system model from physical sensors to remote sensors. Model-based techniques rely on the system model obtained offline; they are very sensitive to changing in the characteristics of the environment and noise. Non-model-based techniques do not use a model of the acoustic plant. Alternatively, they use characteristics of measured signal in order to estimate the error signal in a location of interest that is not physically accessible [5].

III. Conclusion

There are numerous difficulties in creating efficient quite zones by using ANC. Several techniques have been so far proposed; however, they have their own strengths and shortcomings. In moving error microphone approach, the Doppler Effect occurres with the speed of error microphone, resulting in the unstability of the system. Virtual sensing techniques require more sophesticated algorithms that are computationally inefficient. Remote ANC approach that is the direct combination of ANC adaptive algorithm, and virtual sensing technqiues looks promossing as it can achieve a high attenuation degree by modification of the FxLMS algorithm.

IV. References

[1] Kuo, S.M.; Gan, W.S.; Kalluri, S.: Virtual sensor algorithms for active noise control system, in Proc. 2003 IEEE Int. Symp. On Intelligent Signal Processing and communication Systems, December 2003, 714-719.

[2] Kuo, S.M.; Morgan, D.R.: Active Noise Control Systems- Algorithms and DSP Implementations, Wiley, New York, 1996.

[3] I. Tabatabaei Ardekani and W. H. Abdulla, “An adaptive signal processing system for active control of sound in remote location”, APSIPA 2014.

[4] Malgorzata, “Active noise control with moving error microphone”, Mechanics, Vol. 27 No. 3 2008.

[5] Danielle, Ben, Anthony and Peterson, “A review of Virtual Sensing Algorithms for Active Noise Control”, MDPI, Algorithms 2008.

[6] Omid and Ehsan, “Design and implementation of Dual-Microphone adaptive - active noise cancellation system for De noising speech signal”, Electrical review, ISSN 0033-2097, R. 88 NR 2/2012.

[7] Roure, A.; Albarrazin, A. The remote microphone technique for active noise control. In Proceedings of Active 99, pages 1233-1244, 1999.

[8] I. Tabatabaei Ardekani and W. H. Abdulla, “Remote FxLMS algorithm for Active Control of Sound in remote Location”, APSIPA 2014.

[9] Cazzolato, B. An adaptive LMS virtual microphone. In Proceedings of Active 02, pages 105-116, Southampton, UK, 2002.

[10] Petersen, C.; Fraanje, R.; Cazzolato, B.; Zander, A.; Hansen, C. A Kalman filter approach to virtual sensing for active noise control. Mechanical Systems and Signal Processing 2008, 22(2), 490-508.

Ernst Florens Friedrich Chladni was a German physicist and musician. His most important work, for which he is sometimes labeled the "father of acoustics", included research on vibrating plates and the calculation of the speed of sound for different gases. He also undertook pioneering work in the study of meteorites and so is also regarded by some as the "father of meteoritics".

One of Chladni's best-known achievements was inventing a technique to show the various modes of vibration of a rigid surface. When resonating, a plate or membrane is divided into regions that vibrate in opposite directions, bounded by lines where no vibration occurs (nodal lines). Chladni repeated the pioneering experiments of Robert Hooke who, on July 8, 1680, had observed the nodal patterns associated with the vibrations of glass plates. Hooke ran a violin bow along the edge of a plate covered with flour and saw the nodal patterns emerge.

Chladni's technique, first published in 1787 in his book Entdeckungen über die Theorie des Klanges ("Discoveries in the Theory of Sound"), consisted of drawing a bow over a piece of metal whose surface was lightly covered with sand. The plate was bowed until it reached resonance, when the vibration causes the sand to move and concentrate along the nodal lines where the surface is still, outlining the nodal lines. The patterns formed by these lines are what are now called Chladni figures. Similar nodal patterns can also be found by assembling microscale materials on Faraday waves.

Source: https://en.wikipedia.org/wiki/Ernst_Chladni

7th APSIPA ASC 2015

ASIA-PACIFIC SIGNAL AND INFORMATION PROCESSING ASSOCIATION ANNUAL SUMMIT AND CONFERENCE 2015

DECEMBER 16-19, 2015
HONG KONG

apsipa2015.org

Honorary General Chair
Wan-Chi Siu, Hong Kong Polytechnic University

General Co-Chairs
Kenneth Lam, Hong Kong Polytechnic University
Oscar Au, Hong Kong Univ. of Science & Tech.
Helen Meng, Chinese University of Hong Kong

Technical Program Co-Chairs
Changchun Bao, Beijing Univ. of Technology
Akira Hirabayashi, Ritsumeikan University
Jiwu Huang, Shenzhen University
Gwo Giun Lee, National Cheng Kung University
Daniel Lun, Hong Kong Polytechnic University
Tomoaki Ohtsuki, Keio University
Tomasz M. Rutkowski, University of Tsukuba
Sumei Sun, I2R, A*STAR

Finance Chair
Chris Chan, Hong Kong Polytechnic University

Secretary
Bonnie Law, Hong Kong Polytechnic University

Forum Co-Chairs
Wai-Kuen Cham, Chinese Univ. of Hong Kong
Homer Chen, National Taiwan University
King N. Ngan, Chinese University of Hong Kong
Ming-Ting Sun, University of Washington

Panel Session Co-Chairs
Shing-Chow Chan, University of Hong Kong
Dominic K.C. Ho, University of Missouri
Yo-Sung Ho, Gwangju Inst. of Science & Tech.
Yoshikazu Miyanaga, Hokkaido University

Special Session Co-Chairs
Mrityunjoy Chakraborty, India Inst. of Technology
Yui-Lam Chan, Hong Kong Polytechnic University
Lap-Pui Chau, Nanyang Technological University
Hsueh-Ming Hang, National Chiao-Tung University
Hitoshi Kiya, Tokyo Metropolitan University

Tutorial Co-Chairs
Waleed Abdulla, University of Auckland
Woon-Seng Gan, Nanyang Technological Univ.
Wing-Kuen Ling, Guangdong Univ. of Tech.
Lee Tan, Chinese University of Hong Kong

Registration Co-Chairs
Man-Wai Mak, Hong Kong Polytechnic University
Lai-Man Po, City University of Hong Kong

Publication Chair
Zheru Chi, Hong Kong Polytechnic University

Publicity Co-Chairs
Kiyoharu Aizawa, University of Tokyo
Yui-Lam Chan, Hong Kong Polytechnic Univ.
Hing Cheung So, City University of Hong Kong
Mark Liao, IIS, Academia Sinica
Thomas Fang Zheng, Tsinghua University

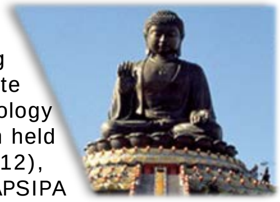
Local Arrangement Co-Chairs
Edward Cheung, Hong Kong Polytechnic Univ.
Frank Leung, Hong Kong Polytechnic University

Advisory Committee

Chairs:
Sadaoki Furui, Toyota Technological Institute at Chicago
Wen Gao, Peking University
C.-C. Jay Kuo, University of Southern California
Haizhou Li, Inst. for Infocomm Research, A*STAR
Ray Liu, University of Maryland

Members:
Thierry Blu, Chinese University of Hong Kong
Shih-Fu Chang, Columbia University
Liang-Gee Chen, National Taiwan University
Li Deng, Microsoft Research
Takeshi Ikenaga, Waseda University
Kebin Jia, Beijing Univ. of Technology
Anthony Kuh, University of Hawaii
Antonio Ortega, University of Southern California
Soo-Chang Pei, National Taiwan University
Susanto Rahardja, National Univ. of Singapore
Yodchanan Wongsawat, Mahidol University
Chung-Hsien Wu, National Cheng Kung Univ.

CALL FOR PAPERS



APSIPA ASC 2015 will be the seventh annual conference organized by the Asia-Pacific Signal and Information Processing Association (APSIPA). Founded in 2009, APSIPA aims to promote research and education in signal processing, information technology and communications. Annual conferences have previously been held in Japan (2009), Singapore (2010), China (2011), the USA (2012), Taiwan (2013), and Cambodia (2014). The field of interest of APSIPA concerns all aspects of signal and information including processing, recognition, classification, communications, networking, computing, system design, security, implementation, and technology with applications to scientific, engineering and social areas. **Accepted papers in regular sessions and special sessions will be published in APSIPA ASC 2015 proceedings which will be submitted for inclusion into IEEE Xplorer as well as other Abstracting and Indexing (A&I) databases.**

The technical program includes the following tracks, but are not limited to:

1. Biomedical Signal Processing and Systems (BioSiPS)
Biomedical Signal and Information: Theory and Methods, Medical Information and Telecare Systems, Neural Systems and Applications, Bio-inspired Signal Processing and System, Biomedical Circuits and Systems.

2. Signal Processing Systems: Design and Implementation (SPS)
Nanoelectronics and Gigascale Systems, VLSI Systems and Applications, Embedded Systems, 3D Video Processing and Coding, High Efficiency Video Coding.

3. Image, Video and Multimedia (IVM)
Image/Video Processing, Coding and Analysis, Image/Video Storage, Retrieval and Authentication, Computer Vision and Graphics, Multimedia Systems and Applications.

4. Speech, Language and Audio (SLA)
Audio Processing, Speech Information Processing: Recognition, Synthesis, Understanding and Translation, Natural Language Processing: Language Modeling, Natural Language Generation/Understanding, Machine Translation.

5. Signal and Information Processing Theory and Methods (SIPTM)
Modern Signal Processing Theory and Method, Detection and Parameter Estimation, Array Processing and Multi-channels, Signal and Information Processing in Applications.

6. Wireless Communications and Networking (WCN)
Information and Network Security, Wireless Communications and Networking, Standards and Emerging Technology, RF and Antennas.

PAPER SUBMISSION

Prospective authors are invited to submit either full papers, up to 10 pages in length, or short papers up to 4 pages in length, where full papers will be for the single-track oral presentation and short papers will be mostly for poster presentation.

IMPORTANT DATES

Submission of Proposals for Special Session, Forum, Panel & Tutorial Sessions:	May 8, 2015
Submission of Full and Short Papers:	June 8, 2015
Submission of Papers in Special Sessions:	June 8, 2015
Notification of Papers Acceptance:	September 1, 2015
Submission of Camera-Ready Papers:	October 1, 2015
Author Registration Deadline:	October 1, 2015
Tutorial Session Date:	December 16, 2015
Summit and Conference Dates:	December 17-19, 2015

IEEE Signal Processing Society

Organizer/Sponsor: The Hong Kong Polytechnic University

Sponsor: Asia-Pacific Signal and Information Association (APSIPA)

Technical Co-Sponsor: IEEE Signal processing Society
IEEE Hong Kong Chapter of Signal Processing

Email: secretary@apsipa2015.org

APSIPA Who's Who

President: Haizhou Li, Institute for Infocomm Research, A*STAR, Singapore

President-Elect: Wan-Chi Siu, The Hong Kong Polytechnic University, Hong Kong

Past President: C.-C. Jay Kuo, University of Southern California, USA

VP - Conferences: Susanto Rahardja, National University of Singapore, Singapore

VP - Industrial Relations and Development: Haohong Wang, TCL Research America, USA

VP - Institutional Relations and Education Program:

Yo-Sung Ho, Gwangju Institute of Science and Technology, Korea

VP - Member Relations and Development: Kenneth Lam, The Hong Kong Polytechnic University, Hong Kong

VP - Publications: Tatsuya Kawahara, Kyoto University, Japan

VP - Technical Activities: Oscar C. Au, Hong Kong University of Science and technology, Hong Kong

Members-at-Large:

Waleed H. Abdulla, The University of Auckland, New Zealand

Kiyoharu Aizawa, The University of Tokyo, Japan

Mrityunjoy Chakraborty, Indian Institute of Technology, India

Kosin Chamnongthai, King Mongkut's University of Technology Thonburi, Thailand

Homer Chen, National Taiwan University, Taiwan

Woon-Seng Gan, Nanyang Technological University, Singapore

Wen Gao, Peking University, China

Hsueh-Ming Hang, National Chiao-Tung University, Taiwan

Anthony Kuh, University of Hawaii at Manoa, USA

K.J. Ray Liu, University of Maryland, USA

Tomoaki Ohtsuki, Keio University, Japan

Ming-Ting Sun, University of Washington, USA

Headquarters

Address:

Asia Pacific Signal and Information Processing Association,
Centre for Signal Processing,
Department of Electronic and Information Engineering,
The Hong Kong Polytechnic University,
Hung Hom, Kowloon, Hong Kong.

Officers:

Director: Wan-Chi Siu, email: enwcsiu@polyu.edu.hk
Manager: Kin-Man Lam, Kenneth,
email: enkmlam@polyu.edu.hk
Secretary: Ngai-Fong Law, Bonnie,
email: ennflaw@polyu.edu.hk
Treasurer: Yuk-Hee Chan, Chris,
email: enyhchan@polyu.edu.hk

Are you an APSIPA member?

If not, then register online at

http://www.apsipa.org