Impact Objectives

- Explore auditory motivated signal processing and the modelling of auditory systems
- Understand more about computational theory of the auditory system using psychology, physiology and information science

An ear for the future

***Professor Masashi Unoki** shares his work investigating acoustics and their use in protecting audio content from piracy and theft*

How did you come to be involved in auditory signal processing and hearing research?

My main research interests are in auditory motivated signal processing and the modelling of auditory systems. I received my MSci and PhD in Information Science from the Japan Advanced Institute of Science and Technology (JAIST) between 1996 and 1999. Here, I focused on acoustical information processing and worked to propose a sound segregation model based on computational auditory scene analysis. I was a Japan Society for the Promotion of Science (JSPS) research fellow from 1998 to 2001. I was also associated with the ATR Human Information Processing Laboratories as a visiting researcher from 1999 to 2000, and I was a visiting research associate at the Centre for the Neural Basis of Hearing (CNBH) in the Department of Physiology at the University of Cambridge from 2000 to 2001. During this period, I worked on constructing a basis for an auditory filterbank as well as auditory signal representation from the results of psychoacoustical and physiological studies. Finally, I have been part of the faculty of the School of Information Science at JAIST since 2001 and am now a full professor.

I feel privileged and very happy that I have had great research experiences and guidance from my current and former supervisors. Professor Masato Akagi was my supervisor in Master/PhD course in JAIST and he is currently a research leader in my group. Dr Roy Patterson was my supervisor in Cambridge University and Professor Brian Moore and Dr Brian Glasberg were also collaborators there. They are true gentlemen and they taught me what scientific research is.

What interests you about your field?

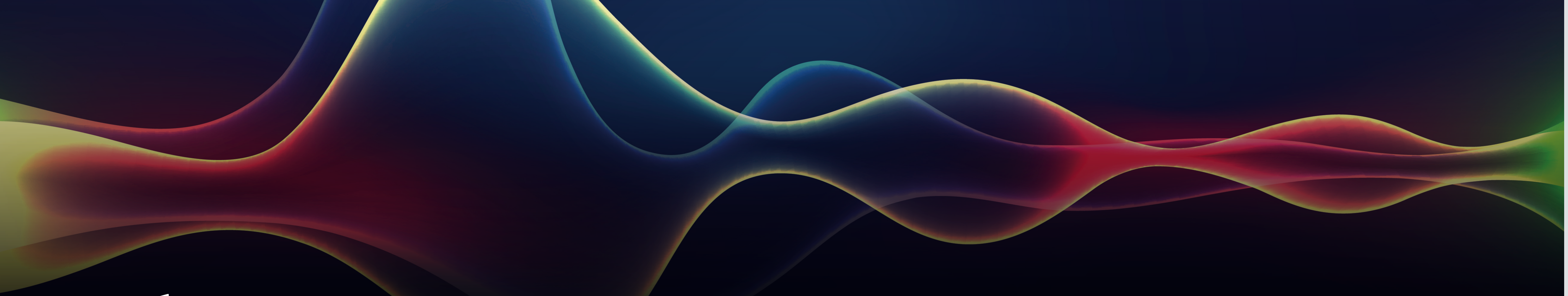
I am very interested in a computational theory of the auditory system. My work tries to answer the questions: what is the purpose of auditory processing? and, why must the auditory system compute it? To do so, we use psychology, physiology and information science. I think it will not only clarify human auditory mechanisms, but also contribute to some applications such as pre-processors for robust speech recognition systems and modelling of the psychoacoustical phenomena. Of course, it should also contribute directly to our research into the use of auditory media signal processing for use against media clone attacks.

What are some of our current knowledge gaps with regards to audio, speech and auditory signal processing?

All these are time sequence signals; however, the signal codec and information compression are quite different between them. For example, MP3/MP4 formats are for music and audio but CELP families are for encoding speech. Although our modality for them is the same, auditory perception, music perception and speech perception are independently studied and developed. There is no good reason, apart from history, for this disparity. There will be various historical and technical reasons. However, I think common signal processing should be established for all these signals. Study on auditory perception will be the common key between them all.

Can you talk about some of the challenges you have faced with this research? How have you overcome these?

There are three challenging issues: satisfying the trade-off between inaudibility for hiding information and robustness against attacks; uncovering unique features for speech fingerprinting; and finding methods for identifying the digital clones from a background signal using the figure-ground segregation concept. We have almost solved the trade-off by studying human auditory perception and using the results from that work. We are still working on uncovering unique features, but my PhD students are currently reconsidering them using space coding techniques such as a spikegram based on auditory sparse coding. Finally, we are still working on how to resolve the identification methods. Our strategy is to understand these issues and to carefully consider them. Never giving up is our spirit for doing our research. ▶



Auditory protection

*Researchers at the **Japan Advanced Institute of Science and Technology** are working on technology that can help watermark and protect digital auditory content*

Understanding how auditory information is processed by the ear and the brain is no easy feat. The mechanics of basic sound signal transduction in the ear are fairly well characterised, however there is much that is still not well-known concerning aspects of the ear and the psychology of the process. Consequently, the way digital sound is stored, processed and delivered fails to take advantage of aspects of human hearing.

Of particular concern is the ease with which digital sound can be virtually copied and distributed without the permission of the creator. Professor Masashi Unoki is an expert in audio/speech signal processing based on auditory perception. He is based at the School of Information Science in the Japan Advanced Institute of Science and Technology (JAIST). He explains that it is extremely difficult to identify illegal copies over the legitimate originals, but that this is not the cause for media, digital or otherwise, that uses the visual senses. Watermarks that are invisible to the human eye but easy to identify if the right kit or knowledge are omnipresent. Banknotes and digital photographs commonly use watermarking to identify fakes.

'Part of the problem is that the human ear is far more sensitive to sound than the eye is to light. It can pick up very quiet sounds as well as a wide range of frequencies,' Unoki points out. 'This makes the ear harder to trick than the eye and therefore it is harder to slip in a piece of sound information that could identify an original from a fake.' Attempting to uncover both more about how humans hear and how this could be used to protect digital sound is the focus of Unoki's latest research efforts. Along with a team of researchers, he is exploring the parts of hearing that are not well understood and hoping to use them for auditory watermarking. 'We are examining aspects of human auditory perception such as masking, cochlear delay and phase perception in order to explore the possibility of auditory information hiding and fingerprinting techniques,' he confirms.

BIOLOGY MEETS TECHNOLOGY

'To be able to discover how and where to hide an auditory watermark, it is necessary to first understand how an ear hears,' Unoki says. Sound waves enter through the ear canal and travel to the eardrum. The eardrum is vibrated by the sound which then moves three very small bones in the middle ear. This amplifies the sound waves and the waves continue travelling to the cochlea. The cochlea is spiral-shaped and filled with uncompressed fluid. The sound travels through this fluid causing disturbances to the two membranes in the cochlear. 'Pressure variations in the cochlear fluids cause the basilar membrane to vibrate in a wavelike motion, different frequency components of a sound cause different parts of the basilar membrane to vibrate,' explains Unoki. 'Information about a sound is represented in the auditory nerve in two ways: in terms of firing rate and in terms of phase locking or firing synchrony.' The sound moves up the auditory nerve and is then processed in the brain. The shape of the cochlea is key to how sound is perceived.

ROBUST EXCHANGE

In order for a digital auditory watermark to be effective, it must display three main characteristics, explains Unoki. 'First, it must be hidden in the sound. If it was placed before or after the piece, it could be easily removed.' Therefore, it must be within the sound itself. Naturally, however it would be extremely undesirable if any extra sound could be heard when played back. Therefore, it must be disguised. 'Secondly, the watermark must maintain confidentiality and accuracy. It must be possible to securely and accurately detect the embedded watermark whenever necessary,' he continues. This is essential to protect the watermark itself and to be able to easily verify originals. 'Finally, the system must be robust. It must be able to withstand changes to the file that occur when transferred and used on different devices and played in different contexts,' says Unoki. Audio files are commonly compressed in order to take up less digital space on portable devices. The watermark would need to be still detectable following such a process.

Unoki and his team are working towards the creation of a watermark that has all three key characteristics. They are using modelling techniques to understand more about the physical mechanisms through which the human ear enables someone to hear. This allows them to understand where in the hearing process additional sounds could be hidden. They have been using a number of modelling tools. 'We mainly use Matlab (matrix laboratories, Mathworks) for our investigations to construct models and methods,' explains Unoki. 'There are many useful toolboxes for analysing data and the implementation of computer programs in various platforms.' In addition, however, the physical side of hearing is only one aspect of the process. The signal for sound is ultimately processed by the brain and there is a strong and important role for the neural and psychological. To this end, Unoki also uses subjective studies with paid volunteers. In these studies, a range of sounds with and without potential watermarks are played to the volunteers who score what they heard based on a series of criteria. This aspect is crucial to validating their modelling. 'It is absolutely essential for us to be able to demonstrate that any auditory watermark does not disrupt the listening enjoyment of the end user,' he states. As such, acoustics research such as Unoki's transcends merely being a branch of physics and incorporates physiology, psychology and computer science too.

> *We are examining aspects of human auditory perception in order to explore the possibility of auditory information hiding and fingerprinting techniques*

THE SOUND OF THE FUTURE

Using this diverse range of approaches, Unoki has developed a watermarking technique based around cochlear delay – the phenomenon of different sound frequencies being transduced in different regions of the cochlea. Essentially, different sounds are heard at different times meaning there is a delay that must be compensated for by the brain. This delay can be exploited by the creation of a sound of the right frequency at the right moment during an audio clip. 'Such a sound would be rendered inaudible by the ear's combination of mechanical and neural compensations that synchronise sounds together,' he notes. 'This sound, however, can be detected by microphones and spotted by computers. This method should fulfil all three criteria of inaudibility, confidentiality and robustness.'

There is still a little way to go, however. Speech recordings and audio files are often stored differently. This means that Unoki's watermarking system needs to be altered to suit the format. They are working hard to adapt the system and Unoki is confident it will succeed. 'The final hurdle is to test the system through making test pirated copies of a file containing the sound,' says Unoki. This is still ongoing, but the progress they have made towards creating a useable watermark to keep creators' content safe has been significant. Not only have they advanced the protection of artists' works, but also they have also helped to shed light on key aspects of the human auditory system and shown that the previous biological studies can be applied directly to useful and impactful technology. ●

The landscape surrounding the JAIST office

Team members: Suradej Duangpumet, Kasorn Galajit, La Pyae Win, Kim Dung Tran, Candy Olivia Mawalim (from left to right)

Project Insights

FUNDING

- Japan Society for the Promotion of Science (JSPS) Grant-in-Aid for Scientific Research (B) Grant Number 17H01761
- Japan Society for the Promotion of Science (JSPS) Grant-in-Aid for Scientific Research (C) Grant Number 18K11609
- Japan Society for the Promotion of Science (JSPS) Grant-in-Aid for Scientific Research Innovative Areas Grant Numbers 16H01669 & 18H05004

COLLABORATORS

- Professor Masato Akagi, JAIST, Japan
- Dr Shengbei Wang, Tianjin Polytechnic University, China
- Dr Weitao Yuan, Tianjin Polytechnic University, China
- Dr Jessada Karnjana, National Science and Technology Development Agency, Thailand

TEAM MEMBERS

- Kim Dung Tran
- Kasorn Galajit
- Suradej Duangpumet
- Candy Olivia Mawalim
- La Pyae Win

CONTACT

Professor Masashi Unoki

T: +81 761 51 1237
E: unoki@jaist.ac.jp
W: http://www.jaist.ac.jp/~unoki/

BIO

Professor Masashi Unoki is based in the School of Information Science, Japan Advanced Institute of Science and Technology. His main research interests are in auditory motivated signal processing and the modelling of auditory systems. Unoki received the Sato Prize from the Acoustical Society of Japan (ASJ) in 1999, 2010, and 2013 for Outstanding Papers and Best Paper Award from the Institute of Electronics, Information and Communication Engineers (IEICE) in 2017.