東北大学 電気通信研究所

研究室外部評価資料

(2013 年度-2018 年度)

Activity Report of Research Laboratory for External Review

April 2013 - March 2019
(FY.2013-2018)

Research Institute of Electrical Communication
Tohoku University

先端音情報システム研究室

Advanced Acoustic Information Systems

April 2013 - March 2019 (FY. 2013-2018)

A. 研究室名 / Research Laboratory

先端音情報システム研究室
Advanced Acoustic Information Systems

B. 構成員 / Faculty and Research Staff (as of May 1, 2019)

※ 欄を適宜追加削除等調整して下さい．期間内に異動等があった場合には，在籍期間を記載して下さい．

教授 / Professor	
氏名 Name	鈴木 陽一 SUZUKI Yôiti (August 1999 -)
分野名 Research Field	先端音情報システム研究分野 Advanced Acoustic Information Systems
准教授 / Associate Professor	
氏名 Name	坂本 修一 SAKAMOTO Shuichi (July 2011 -)
分野名 Research Field	聴覚・複合感覚情報システム研究分野 Auditory and Multisensory Information Systems
助教 / Assistant Professor	
氏名 / Name	大谷 智子 / OHTANI Tomoko (April 2013 - March 2015) トレビーニョ ホルヘ / TREVINO Jorge (February 2015 -) 崔 正烈 / CUI Zhenglie (October 2009 -)
他 / Others	
	産学官連携研究員（特任助教（研究））: 1 名 (October 2016 - March 2019)

C. 研究目的 / Research Purpose

先端音情報システム研究室は，聴覚系及びマルチモーダル知覚情報処理過程に関する基礎研究と，その知見を用いて高度な音響通信システムや快適な音環境を実現するための研究，更にはシステム実現の基礎となるディジタル信号処理の研究に取り組んでいる。これらの研究は，音響学・情報科学だけでなく，電気・通信・電子工学，さらには機械工学・建築学など工学のさまざまな分野や，医学・生理学・心理学などの他の分野とも接点を有する領域にまたがる学際的な性格を特徴としている。

We aim to develop advanced and comfortable acoustic communication systems exploiting digital signal processing techniques. To realize this, we are keenly studying the information processing that takes place in the human auditory system. Moreover, we also investigate the mechanisms for multimodal information processing, including hearing. We mainly apply a psycho-acoustical approach to study human auditory and multimodal perception. In this sense, our research is characterized by its high interdisciplinary nature which covers acoustics, information science, communications engineering, electronics, audiology and psychology.

D. 主な研究テーマ / Research Topics

1. ヒトの聴覚系における情報処理過程の研究
2. ディジタル音信号のセキュアなネットワーク通信に関する研究
3. 3 次元音空間情報の解析，通信及び音像定位制御法の研究
4. 次世代補聴処理システムの研究
5. 環境音の計測・評価・予測法の研究

1. Information processing phenomena in the human auditory system
2. Secure network communications for digital audio signals
3. Analysis, transmission and control of 3D spatial sound information
4. Signal processing for next-generation hearing aids
5. Measurement, evaluation and prediction of ambient sound

E. 学術論文等の編数 / The Number of Research Papers							
	2013	2014	2015	2016	2017	2018	Total
(1) 査読付学術論文 Refereed journal papers	1	10	7	4	9	7	38
(2) 原著論文と同等に扱う 査読付国際会議発表論文 Full papers in refereed conference proceedings equivalent to journal papers	1	0	0	1	1	1	4
(3) 査読付国際会議 Papers in refereed conference proceedings	4	5	2	7	5	9	32
(4) 査読なし国際会議・シンポジウム等 Papers in conference proceedings	5	5	2	7	4	2	25
(5) 総説・解説 Review articles	0	1	0	2	1	5	9
(6) 査読付国内会議 Refereed proceedings in domestic conferences	0	0	0	0	0	0	0
(7) 査読なし国内研究会・講演会 Proceedings in domestic conferences	20	27	18	24	18	25	132
(8) 著書 Books	0	0	1	1	0	2	4
(9) 特許 Patents	0	0	0	2	0	0	2
(10) 招待講演 Invited Talks	11	5	2	6	5	8	37

F. 特筆すべき研究成果 / Significant Research Achievements (FY.2013-2018)

See Ref. 1. “#” mark indicates research carried out at a former organization.

2013-2018 年度の研究成果（論文・特許など）のうち，前半（2013-2015 年度）と後半（2016-2018 年度）それぞれで代表的な数件（2-3 件程度ずつ）について，参考資料を引用して，その特徴と学術的意義などを簡単に紹介する．英文のみ，もしくは和文と英文で記載．
要約は 300 字程度．論文誌の要約/Abstract のコピー可．学術面での国際的インパクトならびに社会的影響を 100 字程度で記載．
必ずしも当該期間内に発表・出版したものに限るのではなく，例えば過去に発表したものでもこの期間内に成果が得られたり，評価されるようになったりしたものも含むものとする．
インパクトファクターや被引用件数など，できる限り第三者が定量的に評価できる指標を用いてアピールすること．それらの指標にはそぐわない場合には，その事情とそれに変わる適当な評価指標・尺度を示すこと．

[2013-2015]

1. S. Sakamoto, G. Hasegawa, A. Honda, Y. Iwaya, Y. Suzuki and J. Gyoba, "Body vibration effects on perceived reality with multi-modal contents," ITE Transactions on Media Technology and Applications, 2(1), pp. 46-50, 2014 [SJR H index: 9][Times Cited: 7]

   Abstract: To develop advanced multi-modal displays, it is important that various sensory information is presented with proper amount. Previous studies have pointed out that vibration information enhances participants' immersive experience from virtual display. Therefore, vibration information should be considered when people develop advanced multimedia systems. We have focused on the effect of full-body vibration information to perceived reality. This study examined the relation between full-body vibration amplitude and perceived reality from audio-visual contents. The sense of presence, the sense of verisimilitude, and the sense of ultra-reality were used as indicators of perceived reality. Results revealed that perceived reality increased by adding full-body vibration. Moreover, results showed that the senses of presence and ultra-reality increased monotonically according to the full-body vibration amplitude, while the sense of verisimilitude exhibited a saturating curvilinear tendency. These results suggest that body vibration is important to increase perceived reality from multi-modal contents presented by advanced multi-modal displays.

   International impact on both academic and social aspects: This work deals with two important metrics to evaluate human perception in multimedia systems: sense-of-presence and verisimilitude. Through this research it was discovered that, when full-body vibration is concerned, the two metrics follow different trends. In particular, the sense of presence increases monotonically with amplitude while the sense of verisimilitude saturates at a certain level. Understanding of this phenomenology is crucial in the development of ultra-realistic communications, tele-presence and virtual reality systems. The results have been well-received in academic circles and are expected to help in the implementation of the next-generation of multimedia interfaces as virtual and augmented reality becomes accessible to end users.

2. S. Sakamoto, S. Hongo, T. Okamoto, Y. Iwaya and Y. Suzuki, "Sound-space recording and binaural presentation system based on a 252-channel microphone array," Acoustical Science and Technology, 36(6), pp. 516-526, 2015. [57th Sato Prize Paper Award, Acoustical Society of Japan], [SJR H index: 33], [Times Cited: 16]

   Abstract: Sensing of high-definition three-dimensional (3D) sound-space information is of crucial importance for realizing total 3D spatial sound technology. We have proposed a sensing method for 3D sound-space information using symmetrically and densely arranged microphones. This method is called SENZI (Symmetrical object with ENchased Zillion microphones). In the SENZI method, signals recorded by the microphones are simply weighted and summed to synthesize a listener’s head-related transfer functions

(HRTFs), reflecting the direction in which the listener is facing even after recording. The SENZI method is being developed as a real-time system using a spherical microphone array and field-programmable gate arrays (FPGAs). In the SENZI system, 252 electric condenser microphones (ECMs) were almost uniformly distributed on a rigid sphere. The deviations of the microphone frequency responses were compensated for using the transfer function of the rigid sphere. To avoid the degradation of the accuracy of the synthesized sound space by microphone internal noise, particularly in the low-frequency region, we analyzed the effect of the signal-to-noise ratio (SNR) of microphones on the accuracy of synthesized sound-space information by controlling condition numbers of matrix constructed from transfer functions. On the basis of the results of these analyses, a compact SENZI system was implemented. Results of experiments indicated that 3D sound-space information was well expressed using the system.

International impact on both academic and social aspects: This work considers the problem of transmitting or preserving spatial sound information in a way that allows the end user to listen to it naturally. The technologies introduced through this research have diverse applications, from tele-presence and virtual reality experiences to the archiving of soundscapes. In the future, it is expected that methods such as SENZI will allow the users to experience events taking place far from their location, or which were archived in the past, while maintaining a high level of realism. In academic circles, the work was recognized by the distinguished Sato Prize Paper Award bestowed by the Acoustical Society of Japan.

[2016-2018]

1. S. Hu, J. Trevino, C. Salvador, S. Sakamoto, J. Li and Y. Suzuki, "A local representation of the head-related transfer function," Journal of the Acoustical Society of America, 140(3), pp. EL285-EL290, 2016 [Impact factor: 1.819][Times Cited: 3]

Abstract: Spatial descriptions of the head-related transfer function (HRTF) using spherical harmonics, which is commonly used for the purpose, consider all directions simultaneously. However, in perceptual studies, it is necessary to model HRTFs with different angular resolutions at different directions. To this end, an alternative spatial representation of the HRTF, based on local analysis functions, is introduced. The proposal is shown to have the potential to describe the local features of the HRTF. This is verified by comparing the reconstruction error achieved by the proposal to that of the spherical harmonic decomposition when reconstructing the HRTF inside a spherical cap.

International impact on both academic and social aspects: This work brings results from disciplines such as computer graphics and Earth sciences and applies them to solve an acoustics problem. In particular, it notices that the commonly used HRTF can be understood as a function on the sphere, similar to lightmaps or topological and gravitational data. The result is a new way to efficiently encode the HRTF; the novel proposal was well received and promptly published by a prestigious journal in the field of acoustics. The methods introduced are expected to be useful to bring realistic 3D sound to portable devices, where storing full HRTF datasets may be difficult.

2. C. Salvador, S. Sakamoto, J. Trevino and Y. Suzuki, "Boundary matching filters for spherical microphone and loudspeaker arrays," IEEE Transactions on Audio, Speech and Language Processing, 26(3), pp. 461-474, 2018. [Impact factor: 3.531], [Times Cited: 3]

Abstract: Conversion of microphone array signals into loudspeaker array signals is an essential process in high-definition spatial audio. This paper presents the theory of boundary matching filters (BMFs) for spherical array signal conversion. BMFs adapt the physical boundary conditions used during recording to the ones required for reproduction by relying on a theoretical framework provided by the Kirchhoff–Helmholtz integral equation (KHIE). Computationally, array signal conversion is performed in a transform domain where sound fields are represented in terms of spherical harmonic functions. Related research on transform-domain signal conversion filters is interpreted in the context of the KHIE. The case of a rigid recording boundary and an open reproduction boundary is addressed. The proposed rigid-to-open BMFs provide a suitable basis for designing gain-limited filters to deal with the problem of excessive gains at certain frequency bands, observed when using high-resolution arrays. Spatial, spectral, and temporal effects in sound field reconstruction when finite numbers of transducers are used in anechoic conditions are investigated analytically and exemplified numerically. Results show that the proposed gain-limited rigid-to-open BMFs outperform the existing gain-limited filters based on Tikhonov regularization because they reduce the spatial discretization effects and yield impulse responses that are more localized around their main peaks.

International impact on both academic and social aspects: This work reviews the state-of-the-art in sound recording and reproduction and makes new proposals to improve the performance in terms of system stability and sound quality. It sheds new light on a very actively researched topic in acoustics which is expected to revolutionize 3D sound for virtual reality and similar applications. The publication above, resulting from this work, gives a full introduction to the field before identifying the issues in classic approaches and specifying how to overcome them. Therefore, it is believed to be useful to scientist who are just taking up research in sound field control.

3. S. Sakamoto, T. Miyashita, Z. Cui, M. Morimoto, Y. Suzuki and H. Sato, "Effects of inter-word pauses on speech intelligibility under long-path echo conditions," Applied Acoustics, 140, pp. 263-274, 2018. [Impact factor: 2.297], [Times Cited: 1]

Abstract: Long-path echo is a salient factor that causes the degradation of the intelligibility of speech transmitted through a wide area outdoor environment or a very large indoor space using public-address systems. To robustly transmit speech information under such conditions, it is important to overcome this effect by controlling the characteristics of speech sounds. In this study, we consider the effects of inserting pauses between the words of a sentence. We performed word intelligibility tests using a series of four continuous words, called a quadruplet. Various pause lengths and long-path echo patterns were applied to the quadruplet. The results of the experiments demonstrate that word intelligibility under a long-path echo is significantly improved by the insertion of pauses between the words. Intelligibility can approach the same levels observed in the absence of echoes for a pause length of approximately 200 ms, which is almost the same as the length of 1-mora for the words used in the experiments. Moreover, this 200 ms pause is known to be sufficient to improve speech recognition in older adults. These results suggest that inter-word pauses of a length of approximately 1-mora can generally enhance the robustness of speech communication systems when used under a severe environment.

International impact on both academic and social aspects: This work focuses on speech intelligibility in conditions that, while uncommon in daily life, are ubiquitous in public address systems such as those used to alert the population of imminent danger. For this reason, research of this kind is thought to have a very high social impact. The results of this work provide a guideline for the speaker to address a population in long path echo conditions while preserving the intelligibility necessary to convey the intended message. The problems tackled in this research have also been very actively discussed in academic circles, particularly in Japan after the Tohoku earthquake and tsunami of 2011.

G. 特筆すべき活動 / Significant Activities (FY.2013-2018)

See Ref. 2-9. "#" mark indicates research carried out at a former organization.

研究室外部評価参考資料の 2 以降を参照しながら，2013-2018 年度のなどの活動の中から特筆すべきものを取り出し，前半（2013-2015 年度）と後半（2016-2018 年度）に分けて簡単に紹介する．英文のみ，もしくは和文と英文で記載．

[2013-2015]

Development of high-definition 3D sound information sensing systems and 3D auditory displays

(supported by Grant-in-Aid for Scientific Research (A) between 2012 and 2015)

This study focused on the development of basic technologies to sense and display 3D spatial sound information. In parallel, it sought to improve our understanding of information processing phenomena taking place in the auditory system and its integration with other perceptual modalities. This understanding is crucial when designing advanced communications technologies and multimedia systems offering a high sense of presence.

In terms of sensing technologies, research focused on the application of spherical and cylindrical microphone arrays. In this context, it produced several important results, such as a new way to bring together human perception features (specifically the head-related transfer function, HRTF) and recording technologies using arrays of dozens or even hundreds of transducers.

As for auditory display technologies, the project succeeded in advancing new applications that traditionally would have been considered too complicated to tackle. As an example, the project achieved the high-definition reproduction of 3D spatial sound in multiple listening zones using a single loudspeaker array.

Through this research, a new perceptual phenomenon was discovered. It was found that the precision of human spatial hearing deteriorates if the listener moves either actively or passively. This mechanism was modeled, and the results are expected to be useful in the development of future high-definition 3D sound systems.

The output of this project includes a total of 21 publications in scientific journals and over 40 presentations in conferences.

[2016-2018]

Development of free-listening-point 3D sound systems based on ADVISE theory

(supported by Grant-in-Aid for Scientific Research (A) between 2016 and 2019)

This research project aimed to develop high-definition, 3D sound systems that account for the perceptual phenomenology that was identified during the project described above. In particular, the need to support what is referred to as human active listening, i.e. phenomena involving auditory perception while the listener is in motion. This called for the proposal of new technologies that can control sound at a freely placed listening position. Noticing this requirement, the project makes use of the theory of 3D auditory displays based on the so-called virtual sphere model (ADVISE). In particular, the problem of developing a free-listening-point 3D sound system was broken into three parts: a room acoustics problem which handles the environmental information, an auditory perception problem which produces the required sound signals to convey 3D sound, and the development of an interface capable of matching these two.

During the course of this investigation, some limitations of the classical theory of ADVISE were identified and the pertinent modifications were made to overcome all potential issues in its implementation. Recent results in spatial sound control, such as the proposal of so-called boundary matching filters, helped to finalize a new formulation of ADVISE that is free of the problems observed in the original proposal, while making the task of freely displacing the listening position easier to implement.

In addition, the project advanced a modified version of a room acoustics model called Adaptive Rectangular Decomposition (ARD). The proposal replaces the room modes used in ARD with plane waves. An explicit translation operator for these plane waves was introduced, making it easy to change listening positions without having to recalculate the room acoustics model, a computationally expensive task.

The results described above allowed for the implementation of a 3D auditory display based on the revised version of ADVISE. Further, the project complemented said result by also developing new high-definition efficient systems for the acquisition of 3D spatial sound. The focus was on the use of compact, spherical microphone arrays, including arrays of arrays (i.e. multiple spherical microphone arrays used together as a single system). Finally, the project also involve new research into the perceptual aspects of hearing in active listening conditions; these results helped focus the research efforts throughout the project.

The output of this project includes a total of 10 journal papers published thus far, as well as over 30 presentations in academic conferences.