# Project Title: Spoof Detection for Automatic Speaker Verification

**Background :** The human voice is one of the biometric indicators representing a person's identity. It has been successfully used in automatic speaker verification (ASV) systems in many applications. The ASV system is currently vulnerable to spoofing attacks in which someone disguises as another and illegitimately accesses a secure system. Hence, countermeasures against spoofing attacks are necessary to verify whether the claimed voice is a genuine or fake representation before verifying or identifying the speaker's identity. Therefore, this research focuses on spoofing detection in automatic speaker verification.

**Targets:** The objectives of this project are listed as follows:
(1) To explore and investigate significant of speech features for spoof detection,
(2) To optimize percentage of voice and non-voice segments in features used in spoofing detection,
(3) To investigate pathological feature for spoof detection
(4) To minimize detection error,
(5) To improve an accuracy of ASV, and
(6) To study multi-lingual spoof detection.



**Speaker:** Kasorn GALAJIT(Project Leader and Speaker)
National Electronics and Computer Technology Center, Thailand
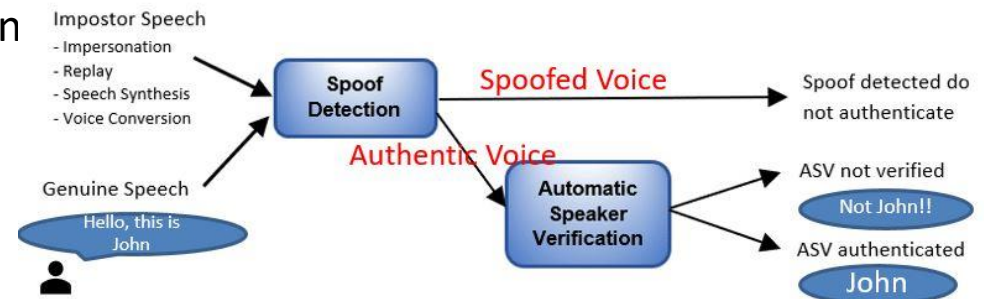
# Project Title:

**Project Members :**  25 members and 3 pending members

| Name / Position / Institution | Name / Position / Institution | Name / Position / Institution | Name / Position / Institution | Name / Position / Institution |
|---|---|---|---|---|
| Masashi Unoki (PhD) / Professor / JAIST, Japan | Pakinee Aimmanee / Assoc. Prof. / SIIT, Thailand | Kasorn Galajit (PhD) / Project Leader / NECTEC, Thailand | Win Lai Lai Phyu / Asst. Lecturer / UCSY, Myanmar | Waree Kongprawechnon (PhD) / Assoc. Prof. / SIIT, Thailand |
| Candy Olivia Mawalin (PhD) / Asst. Prof. / JAIST, Japan | Khaing Zar Mon / Grad student / SIIT, Thailand | Suradej Duangpummet (PhD) / Researcher / NECTEC, Thailand | Myat Aye Aye Aung / Asst. Lecturer / UCSY, Myanmar | Patthranit Kaewcharuay / Undergrad student / SIIT, Thailand |
| Kai Li / PhD Candidate / JAIST, Japan | Ananda Garin Mills / Undergrad student/ SIIT, Thailand | Jessada Karnjana (PhD) / Researcher / NECTEC, Thailand | Win Pa Pa (PhD) / Professor / UCSY, Myanmar | Nanthayod Termkoh / Undergrad student / SIIT, Thailand |
| Anuwat Chaiwongyen / PhD Candidate / JAIST, Japan & SIIT, Thailand | Pannathorn Sathirasattayanon / Undergrad student / SIIT, Thailand | Widhyakorn Asdornwised (PhD) / Asst. Prof. / CU, Thailand | Aye Mya Hlaing (PhD) / Lecturer / UCSY, Myanmar | Xugang Lu (PhD) / Senior Researcher / NICT, Japan |
| Dessi Puji Lestari (PhD) / Professor / ITB, Indonesia | Dk Hayati Pg Hj Mohd Yassin (PhD) / Lecturer / UBD, Brunei | Navod Neranjan Thilakarathne / PhD Candidate / UBD, Brunei | Hay Mar Soe Naing/ Lecturer/ UCSY, Myanmar | Sheng Li (PhD) / Senior Researcher / NICT, Japan |

**Project Duration :**  2 years, 1 April 2023 – 31 March 2025

| Name / Position / Institution (Pending) |
|---|
| Sasiporn Usanavain(PhD) / Assoc. Prof. / SIIT, Thailand |
| Surasak Boonkla (PhD) / Researcher / NECTEC, Thailand |
| Kosin  Kalarat / Undergrad student / SIIT, Thailand |

**Project Budget:**  80,000 USD  (40,000 USD per year)

# Project Activities: Scientific

- NECTEC, JAIST, SIIT, USCY, UBD Provides "ThaiSpoof" Dataset for spoof detection for Thai language
  Researchers can use this dataset to study spoof detection
- JAIST and NICT study the "Contributions of Jitter and Shimmer in the Voice for Fake Audio Detection"
  Statistical analysis results indicate prosody differences captured by the shimmer features, especially the CS3, can provide important information to distinguish between fake and genuine speech. (Journal)
- SIIT, JAIST, NECTEC Study Spoof Detection using Voice Contribution on LFCC features and ResNet-34
  Difference percentage of voice contribute to accuracy to detect spoofing

# Project Activities: Scientific

- **UBD,** NECTEC, JAIST, UCSY study Using Novel Hybrid Convolutional Neural Network for Dysarthria Diagnosis
  This study explorer the characteristic of abnormal voice.

- **UCSY** study A Large Vocabulary End-to-End Myanmar Automatic Speech Recognition
  The study gives knowledge of speech recognition in Myanmar language

- **JAIST,** NECTEC, SIIT study Deepfake-speech detection with pathological features and multilayer perceptron neural network
  This study can show the significant of pathological for spoof detection

- **UCSY** study  M-Diarization: A Myanmar Speaker Diarization using Multi-scale dynamic weights
  - The study gives knowledge of speech recognition in Myanmar language, Also introduce the Myanmar dataset for speaker recognition as well as diarization.

- **NECTEC ,** SIIT study Speech Watermarking for Tampering Detection Using Singular Spectrum Analysis with a Psychoacoustic Model
    - This study can show characteristic of tampering speech which can be used for spoof detection

**Budget**: 40,000 USD/year

| | |
|---|---|
| Meeting/workshop (2 times a year) | 14,000 USD |
| IVO Forum | 1,000 USD |
| Publication, presentation at international conferences | 25,000 USD |

## Meeting (14,000 USD)

| May | Jun | Jul | Aug | Sep | Oct | Nov | Dec | Jan | Feb | Mar | Apr |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Kick-off | | | | | 2nd | Forum | | | | | |
| 3rd | | | | | | Forum | Final | | | END | |

Kick-off meeting **2,500**USD (Thailand May 2023) ,
Second meeting  **6,000**USD (Indonesia Dec 2023)

### Publication

| | |
|---|---|
| ISAI Thailand  2 papers | 800USD |
| SCDE Fiji1 Paper | 2,700USD |
| Cocosda India 2 papers | 2,1260USD |
| ACM 1 paper | 310USD |
| APSIPA2023  1 paper | 0 USD |
| IEEE Access 1 Journal | 0 USD |
| Total **5970** USD | |

## Conference/Publication (25,000 USD)

- Total 50,000 USD for 2 years
  - 2,500 USD (85,000 THB or 5,250,000 Kyat or 3,300 BND approx.) each

| | |
|---|---|
| NECTEC | 3 |
| SIIT | 2 |
| CU | 1 |
| UCSY | 2 |
| UBD | 1 |
| | 9 |

**JAIST
NICT**

Publication fee
Conference registration

### Approximately used budget

| | |
|---|---|
| Meeting | 8,500 |
| Publication | 5,970 |
| Forum | 1,000 |

Total **15,400** USD

# R&D results:

- "ThaiSpoof" Dataset for spoof detection for Thai language

   This dataset will be put in AI for Thai for sharing
-  Myanmar dataset for speaker recognition and diarization.
- VAJA

   Thai speech synthesis in AI for Thai
- The shimmer features, especially the CS3, can provide important information to distinguish between fake and genuine speech.
- Difference percentage of voice contribute to accuracy to detect spoofing
- The characteristic of abnormal voice.
- Speech recognition for Myanmar language
- The significant of pathological for spoof detection
- The characteristic of tampering speech which can be used for spoof detection

# Scientific Contribution:

Presentations at International Conferences:

**Expected output in 2 years**
10 Conference papers

| No: | Paper title: | Author names | Affiliation | Conference name: | The date of the conference | The venue of the conference |
|---|---|---|---|---|---|---|
| 1 | ThaiSpoof: A Database for Spoof Detection in Thai Language | • Kasorn Galajit, <br>• Thunpisit Kosolsriwiwat, <br>• Candy Olivia Mawalim, <br>• Pakinee Aimmanee, <br>• Waree Kongprawechnon, <br>• Win Pa Pa, <br>• Anuwat Chaiwongyen, <br>• Teeradaj Racharak, <br>• Hayati Yassin, <br>• Jessada Karnjana, <br>• Surasak Boonkla, <br>• Masashi Unoki, | • NECTEC, National Science and Technology Development Agency, <br>• Sirindhorn International Institute of Technology, <br>• Japan Advanced Institute of Science and Technology, <br>• University of Computer Studies,Yangon, <br>• Universiti Brunei Darussalam}, Brunei Darussalam | The 18th International Joint Symposium on Artificial Intelligence and Natural Language Processing (iSAI-NLP 2023) (https://isai-nlp-aiot2023.aiat.or.th/) | 27-29/11/2023 | Bangkok, Thailand |
| 2 | Spoof Detection using Voice Contribution on LFCC features and ResNet-34 | • Khaing Zar Mon, <br>• Kasorn Galajit, <br>• Candy Olivia Mawalim, <br>• Jessada Karnjana, <br>• Tsuyoshi Isshiki, <br>• Pakinee Aimmanee, | • Sirindhorn International Institute of Technology, <br>• NECTEC, National Science and Technology Development Agency, <br>• Japan Advanced Institute of Science and Technology, <br>• Tokyo Institute of Technology Tokyo, | The 18th International Joint Symposium on Artificial Intelligence and Natural Language Processing (iSAI-NLP 2023) (https://isai-nlp-aiot2023.aiat.or.th/) | 27-29/11/2023 | Bangkok, Thailand |
| 3 | Using Novel Hybrid Convolutional Neural Network for Dysarthria Diagnosis | • Navod Neranjan Thilakarathne, <br>• Kasorn Galajit, <br>• Jessada Karnjana, <br>• Win Pa Pa, <br>• Candy Olivia Mawali, <br>• Hayati Yassin, | • Universiti Brunei Darussalam}, Brunei Darussalam <br>• NECTEC, National Science and Technology Development Agency, <br>• Japan Advanced Institute of Science and Technology, <br>• University of Computer Studies,Yangon, | The 10th IEEE CSDE 2023, the Asia-Pacific Conference on Computer Science and Data Engineering 2023,(IEEE CSDE 2023) https://ieee-csde.org/csde2023/ | 4-6/12/2023 | Yanuca Island, Fiji |

# Scientific Contribution:

Presentations at International Conferences:

| No: | Paper title: | Author names | Affiliation | Conference name: | The date of the conference | The venue of the conference |
|---|---|---|---|---|---|---|
| 4 | Speech Watermarking for Tampering Detection Using Singular Spectrum Analysis with a Psychoacoustic Model | • Phondanai Khanti,<br>• Pannathorn Sathirasattayanon,<br>• Patthranit Kaewcharuay,<br>• Nanthayod Termkoh,<br>• Ekachai Phaisangittisagul,<br>• Kasorn Galajit,<br>• Jessada Karnjana | • Sirindhorn International Institute of Technology,<br>• NECTEC, National Science and Technology Development Agency,<br>• Kasetsart University | The 26th Conference of the Oriental COCOSDA https://www.ococosda2023.com/ | 4-6/12/2023 | Delhi, India |
| 5 | A Large Vocabulary End-to-End Myanmar Automatic Speech Recognition | • Hay Mar Soe Naing<br>• Win Pa Pa | • University of Computer Studies,Yangon, | M3Oriental Workshop of ACM Multimedia Asia 2023 The ACM Multimedia Asia 2023 https://sites.google.com/view/m3oriental | 8/12/2023 | Tainan city, Taiwan |
| 6 | Deepfake-speech detection with pathological features and multilayer perceptron neural network | • Anuwat Chaiwongyen,<br>• Suradej Duangpummet,<br>• Jessada Karnjana,<br>• Waree Kongprawechnon,<br>• Masashi Unoki | • Japan Advanced Institute of Science and Technology,<br>• NECTEC, National Science and Technology Development Agency,<br>• Sirindhorn International Institute of Technology, | The 15th annual conference organized by Asia-Pacific Signal and Information Processing Association (APSIPA2023) https://www.apsipa2023.org/ | 31/11/2023-3/12/2023 | Teipei, Taiwan |

# Scientific Contribution:

| No: | Paper title: | Author names | Affiliation | Conference name: | The date of the conference | The venue of the conference |
|---|---|---|---|---|---|---|
| 7 | M-Diarization: A Myanmar Speaker Diarization using Multi-scale dynamic weights | Myat Aye Aye Aung, Win Pa Pa, Hay Mar Soe Naing. | • University of Computer Studies,Yangon,, | The 26th Conference of the Oriental COCOSDA https://www.oco cosda2023.com/ | 4-6/12/2023 | Delhi, India |

# Scientific Contribution:

| No: | Paper title: | Author names | Affiliation | Journal name: | The publisher of the Journal | The volume number and Pages |
|-----|-------------|--------------|-------------|---------------|------------------------------|----------------------------|
| 1 | Contributions of Jitter and Shimmer in the Voice for Fake Audio Detection | KAI LI, XUGANG LU, MASATO AKAGI, MASASHI UNOKI. | • Japan Advanced Institute of Science and Technology, • Advanced Speech Technology Laboratory, National Institute of Information and Communications Technology, | IEEE Access | IEEE Access | VOLUME 11, 2023 Received 21 June 2023, accepted 29 July 2023, date of publication 3 August 2023, Digital Object Identifier 10.1109/ACCESS.2023.3301616 |

**Expected output in 2 years**
10 Conference papers
2 Journals

**Current output (approximately 1 year)**
7 Conference papers
1 Journals

# Societal Impact:

- **ThaiSpoof: A Database for Spoof Detection in Thai Language**

A SUMMARY OF THAISPOOF DATABASE FOR SPOOF DETECTION.

| Label | Database type | Degree | No. utterrance |
|-------|---------------|--------|----------------|
| Genuine | Genuine dataset | - | 143,262 |
| Spoofed | Text-to-speech dataset | - | 143,262 |
| | F0 modification dataset | 10 ch/oct | 143,262 |
| | | 40 ch/oct | 143,262 |
| | | 160 ch/oct | 143,262 |
| | | 320 ch/oct | 143,262 |
| | Pitch shifting | +4% | 143,262 |
| | | +10% | 143,262 |
| | | +20% | 143,262 |
| | | -4% | 143,262 |
| | | -10% | 143,262 |
| | | -20% | 143,262 |



M3Oriental @ ACM MMAsia2023
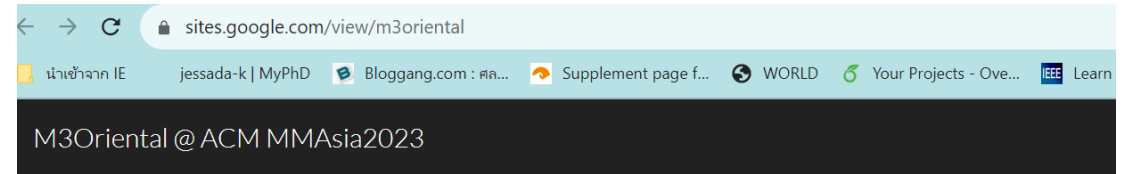
**Organizers and Program Commitee**

This workshop is partly supported by NICT international funding.

For questions, please contact organizers at: sheng.li-a-t-nict.go.jp or other organizers.

**Speech**

Dr. Eng Siong Chng, Nanyang Technological University (NTU), Singapore, Associate Professor (ASESChng-a-t-ntu.edu.sg)

Dr. Zhizheng Wu, Chinese University of Hong Kong, Shenzhen, Associate Professor (wuzhizheng-a-t-cuhk.edu.cn)

Dr. Xugang Lu, NICT, Kyoto, Japan, Senior Researcher (xugang.lu-a-t-nict.go.jp)

Dr. Sheng Li, NICT, Kyoto, Japan, Researcher (sheng.li-a-t-nict.go.jp)

Dr. Xinhui Hu, RoyalFlush AI, China, Chief Scientist (huxinhui-a-t-myhexin.com)

- **Myanmar dataset for speaker recognition and diarization.**
- **NICT project member organizing the M3Oriental @ ACM MMAsia2023**

Conclusion:

- Dataset :"ThaiSpoof" Dataset for  spoof detection for Thai language
- Myanmar dataset for speaker recognition and diarization.
- The knowledge of the contributions of Jitter and Shimmer in the Voice for Fake Audio Detection
- The knowledge of the contributions of Jitter and Shimmer in the Voice for Fake Audio Detection
- The knowledge of the contributions of voice and non-voice for detect spoofing in MFCC and LFCC
- The knowledge of characteristic of abnormal voice.
- The knowledge of speech recognition in Myanmar language
- The Knowledge of the significant of pathological for spoof detection
- The Knowledge of characteristic of tampering speech which can be used for spoof detection

# Future works:

- Improving Dataset :"ThaiSpoof" using new vocoder
- Data fusion of difference feature.
- API development
- Cross dataset and cross language in AMS language