Spoof Detection for Automatic Speaker Verification

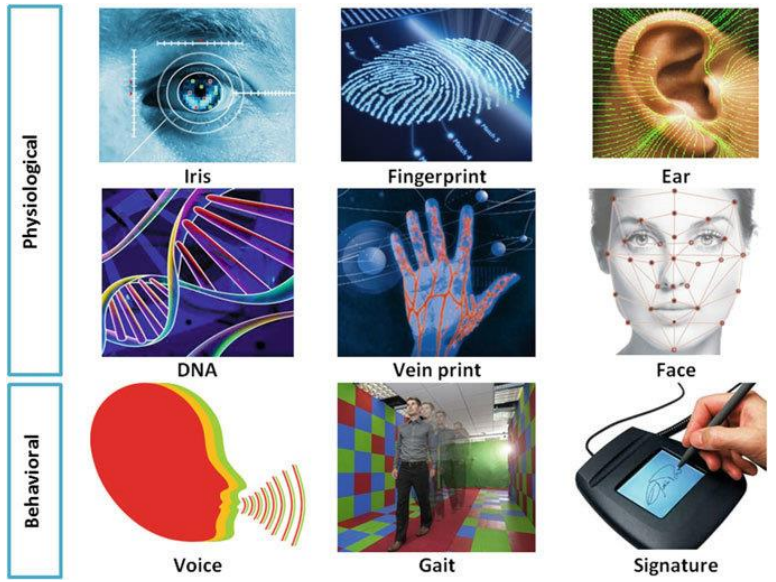
Kasorn GALAJIT

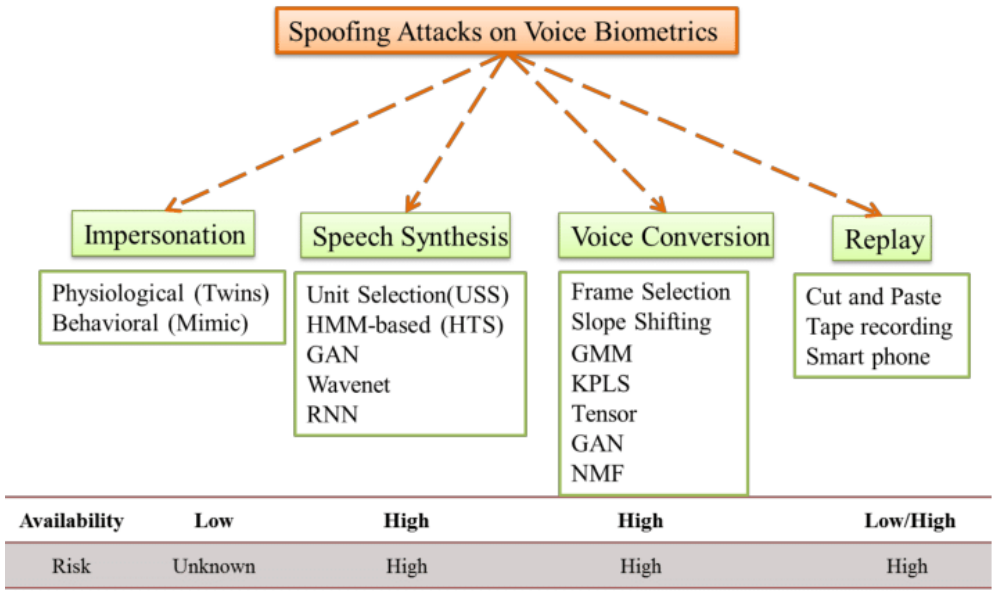
National Science and Technology Development Agency (NSTDA), Thailand
Sirindhorn International Institute of Technology (SIIT), Thailand
Japan Advanced Institute of Science and Technology (JAIST), Japan

Spoof Detection for Automatic Speaker Verification

Background :

- Voice is 15% sharing in applications of biometrics
- Primarily used in technology and financial domains





Problem !! spoofing attacks

Spoofing by using tools/advanced technology without prior knowledge

- Countermeasures against spoofing attacks is necessary
  - to verify whether the claimed voice is a genuine

Image Source: Bouchrika, Imed. "A survey of using biometrics for smart visual surveillance: Gait recognition." Surveillance in Action. Springer, Cham, 2018. 3-23.
Akbar, Muhammad Jalaluddin. "A Overview of Spoof Speech Detection for Automatic Speaker Verification." (2019).

Spoof Detection for Automatic Speaker Verification

ASEAN IVO Network Expansion

Ongoing Project

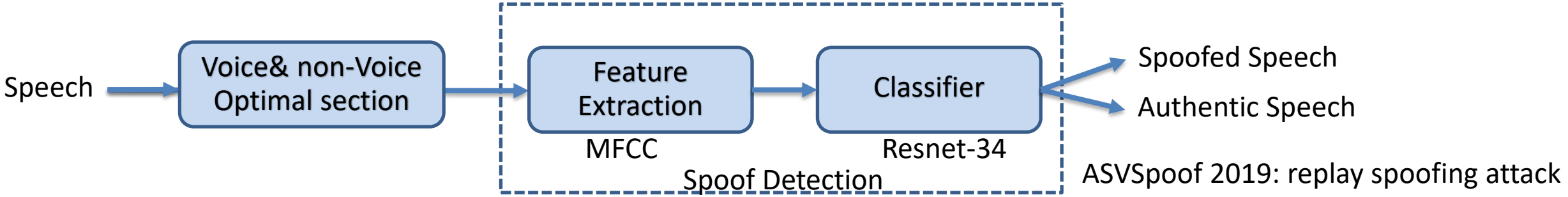
- Contributions of voice & non-voiced for spoof detection
- Pathological features (timbre and shimmer) for deepfake detection

SIIT SINCE 1992 NECTEC a member of NSTDA JAIST JAPAN ADVANCED INSTITUTE OF SCIENCE AND TECHNOLOGY 1990

- Significant of speech features for spoof detection
- Contributions of voice & non-voiced of each feature for spoof detection
- Pathological features for spoof detection
- Minimize error detection
- Multi-lingual spoof detection

NICT

Spoof Detection for Automatic Speaker Verification : Ongoing Project 1



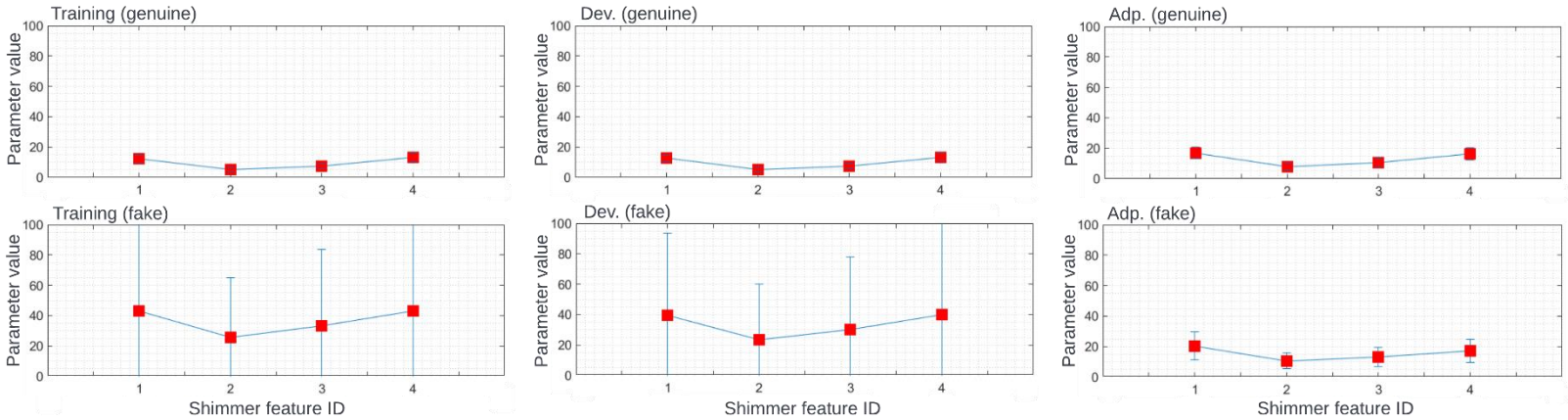
ASVSpoof 2019: replay spoofing attack

The contributions of the proportional voice and non-voiced sections for replay attack detection.

	EER (%)		Accuracy (%)
	Dev	Eval	Eval
Whole utterance	**1.72**	**1.86**	**98.90**
Voice only	32.04	31.45	68.55
Non-voice only	2.05	2.90	97.10

Feature (MFCC)	EER (%)		Accuracy (%)	F0.5 (%)	F1 (%)	F2 (%)	Precision (%)	Recall (%)
	Dev	Eval	Eval					
Whole utterance	1.72	1.86	98.14	99.39	98.91	98.45	99.71	98.14
Non voice + 5% voice	1.63	1.89	98.11	99.38	98.90	98.42	99.70	98.10
Non voice + 10% voice	**1.43**	1.76	98.24	99.42	98.98	98.54	99.72	98.24
Non voice + 20% voice	1.53	**1.72**	**98.28**	**99.44**	**99.00**	**98.56**	**99.73**	**98.28**
Non voice + 30% voice	1.62	1.77	98.22	99.42	98.97	98.52	99.72	98.22
Non voice + 40% voice	1.71	1.77	98.23	99.42	98.97	98.52	99.72	98.22

- Speech synthesis ➔high-quality voices, naturalness is close to human voices.
- a malicious synthesis voice, called audio deepfake ➔cases of fraud.

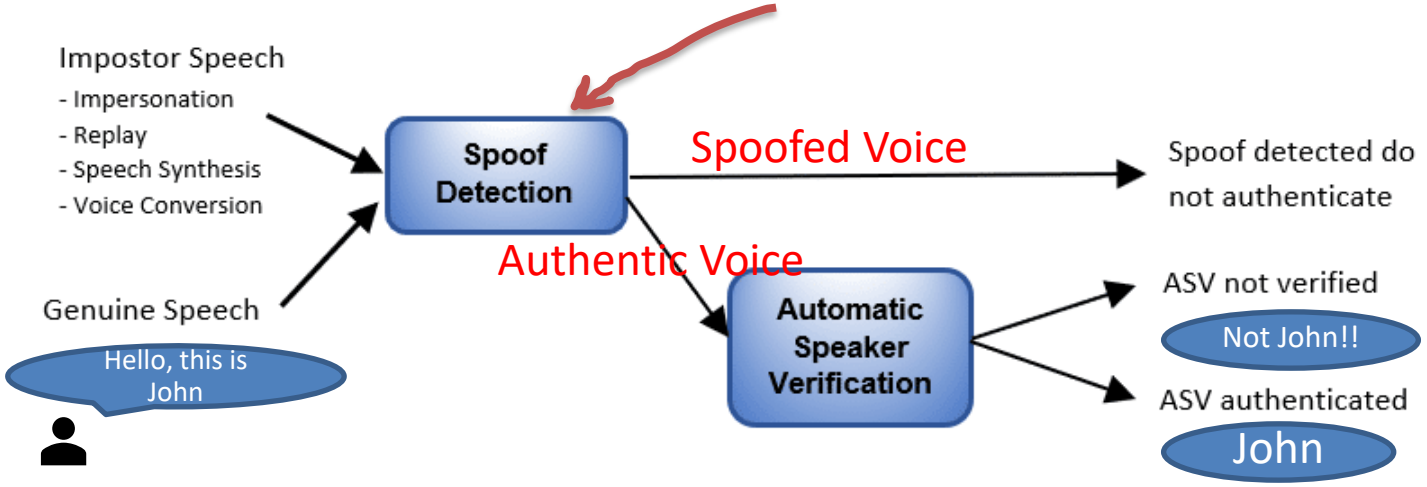


the attributions of timbre and shimmer features have high contributions to detecting deepfake speeches

Pathological voice is analyzed to be the feature for distinguishing human voices and deepfakes.

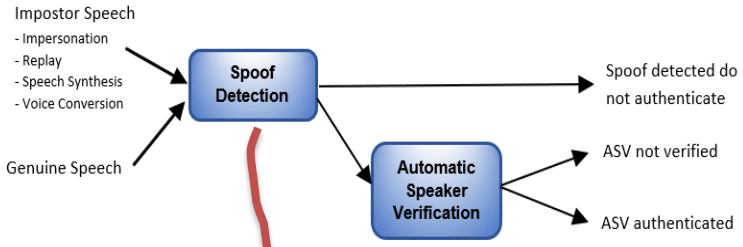
pathological voice is usually used for diseases diagnostic from the quality of voice such as oral cancer and voice disorder.

Proposed Method (idea): Spoof Detection



- To check significant of speech features for spoof detection
- To optimal percentage of voice and non-voice for each feature
- To investigate pathological feature for spoof detection
- To minimize detection error
- To improve an accuracy of ASV
- To study multi-lingual spoof detection

Proposed Method (idea):

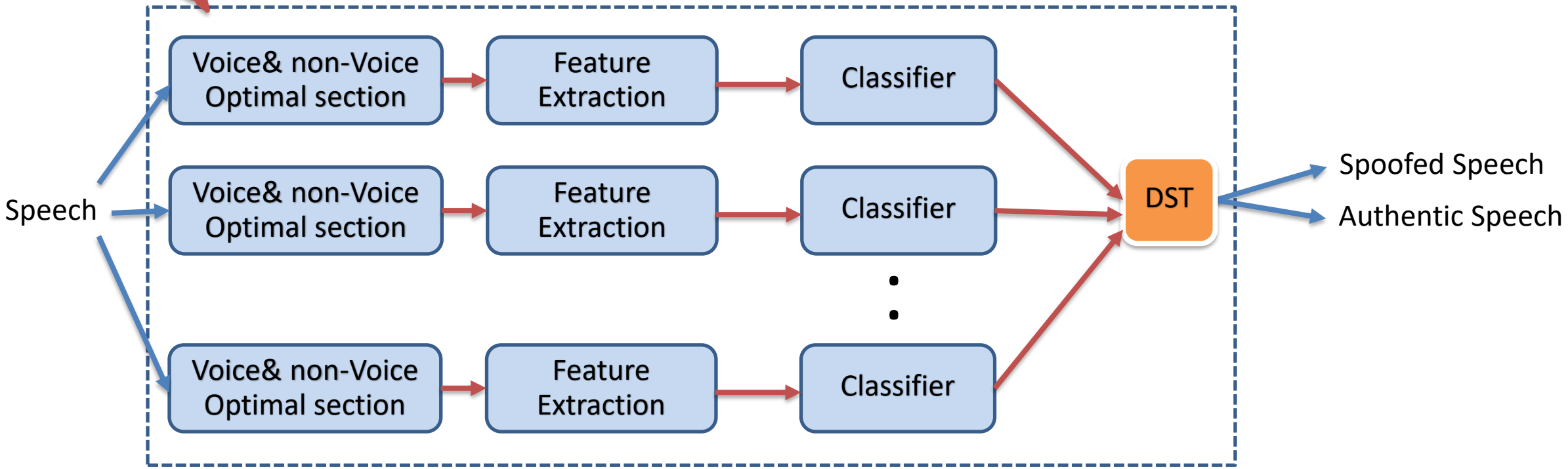


Multi-lingual

Dataset :

ASVspoof
Automatic Speaker Verification and Spoofing Countermeasures Challenge

Common Voice
moz://a



DST : Dempster-Shafer theory, a framework based on a bounded-error estimation method

Impact:

Scientific & Technological Impact

This project's direct scientific/technological impact is knowledge obtained from the exploration and investigation of the use of pathological features for spoofing detection. The detection accuracy can be improved. The number of false positives and negatives can be reduced. Also, techniques relating to data fusion can be applied to many more applications. Thus, this is an indirect impact.

In addition, to develop machine learning models, necessary databases are to be constructed. They will be helpful for further development in the field.

Societal Impact

The progress in biometric authentication and verification in real-world applications has been realized in several mobile banking applications and online shopping platforms. Thus, the security issue is crucial. Consequently, this project contributes directly to cyber security in society.

Output/Outcome

Target:

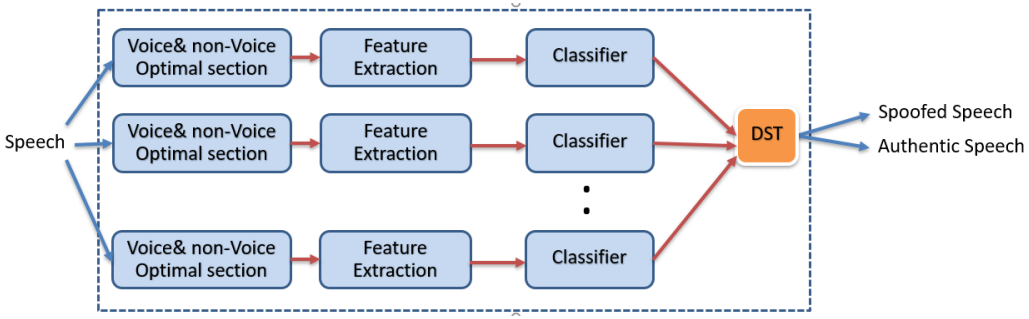
- To check significant of speech features for spoof detection
- To optimal percentage of voice and non-voice for each feature
- To investigate pathological feature for spoof detection
- To minimize detection error
- To improve an accuracy of ASV
- To study multi-lingual spoof detection

- Scientific,
  - Publication
  - Prototype
  - Dataset for spoof detection for ASEAN language
  - Patent
- Societal,
  - API of spoof detection for public used.
- Collaborative,
  - Need new partner for multi-lingual

Conclusion:

Problem : Spoofing attack in Automatic Speaker Verification

1. Targets : Spoof detection
   - To check significant of speech features for spoof detection and optimal percentage of voice and non-voice for each feature
   - To investigate pathological feature for spoof detection
   - To minimize detection error
   - To study multi-lingual spoof detection



2. Method (idea):
   - Significant features for spoof detection
   - Optimal voice and non-voice for each feature
   - Pathological feature
   - Dempster-Shafer theory model to minimize detection error

3. Scientific and societal impact:
   - Publications, Patent, Prototype
   - Dataset for spoofed detection for ASEAN language
   - Corporation for multi-lingual