

# AI Application

Sichen Liu

INT104 Artificial Intelligence



# AI Application

**Smart Home** : smart speaker, smart screen, smart camera, sweeping robot, smart door lock, smart gateway ;

**Smart Wearable device** : smart watch, smart glasses, smart clothes, AI tracker, health medical equipment, AR/VR with AI ;

**Smart City** : self-driving vehicle, smart traffic monitor, smart traffic light ;

**Smart Logistics** : distribution robot, unmanned delivery vehicle, distribution UAV ;

**Smart Retail** : self-service market, shopping guide robot, smart vending machine;

**Smart Community** : intelligent elevator, smart camera ;

**Smart Factory** : sorting robot ;

**Animal Husbandry and Agriculture** : plant protection UAV, self driving tractor ;

**Wise Medical** : medical robot, smart wearable device ;

**Education** : tutor/evaluation robot ;

**Security** : AI monitor;



# AI Application

## Automatic Speech and Semantic Recognition

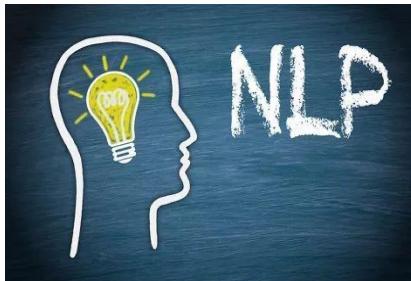
### Smart Voice System



- speech recognition
- speech synthesis
- speech assessment

.....

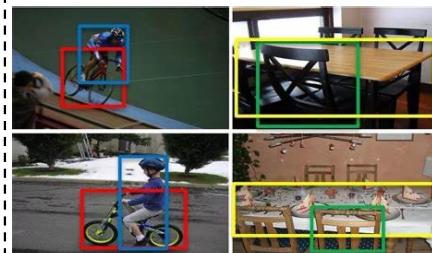
### Natural Language Processing



- language translation
- phrase analysis
- semantic analysis

.....

### Computer Vision



Person & bicycle      chair & table

- image classification
- object detection
- target tracking
- image segmentation

.....

### Biometric Authentication



- fingerprint identification
- face recognition
- vein recognition

.....



# Communication

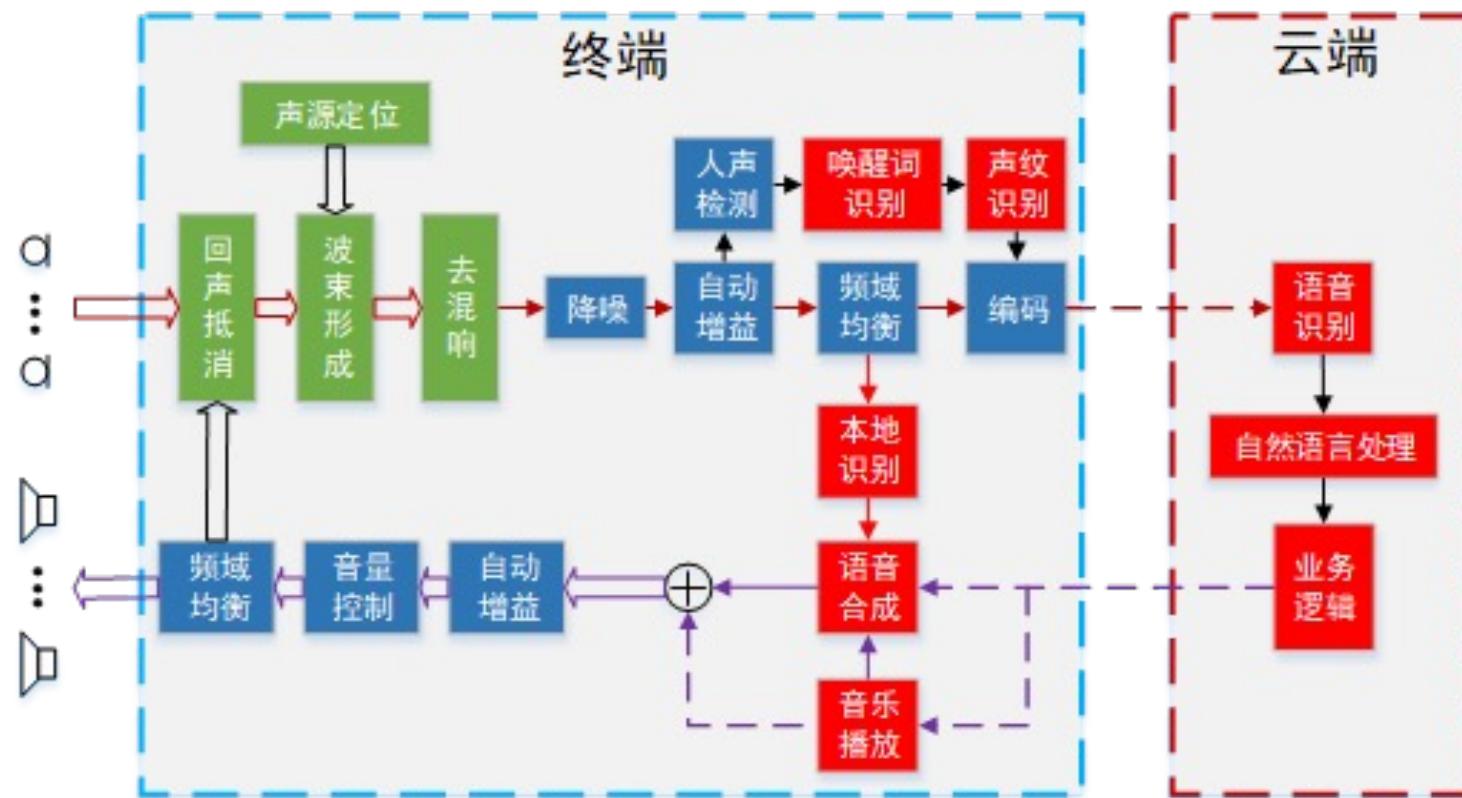
# Acoustics

XJTLU | SCHOOL OF  
FILM AND  
TV ARTS



# Communication Acoustics

## Acoustic Framework of Smart Speaker



# Communication Acoustics

---

## Acoustic Echo Cancellation

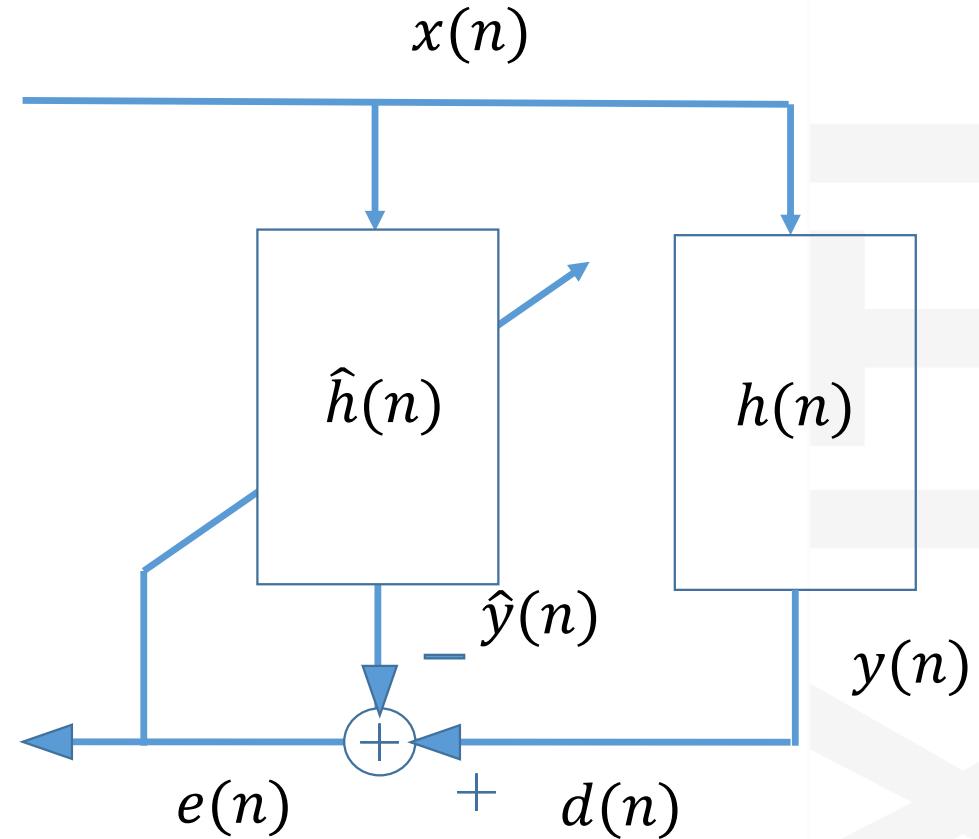
- **Echo from remote side:**
  - The speaker will hear his own voice with delay, which is annoying and confusing.
  - The speech from the remote speaker will be affected by the echo signal.
- **Echo on both sides:**
  - Worst-case scenario. Feedback can make communication impossible.



# Communication Acoustics

## Acoustic Echo Cancellation

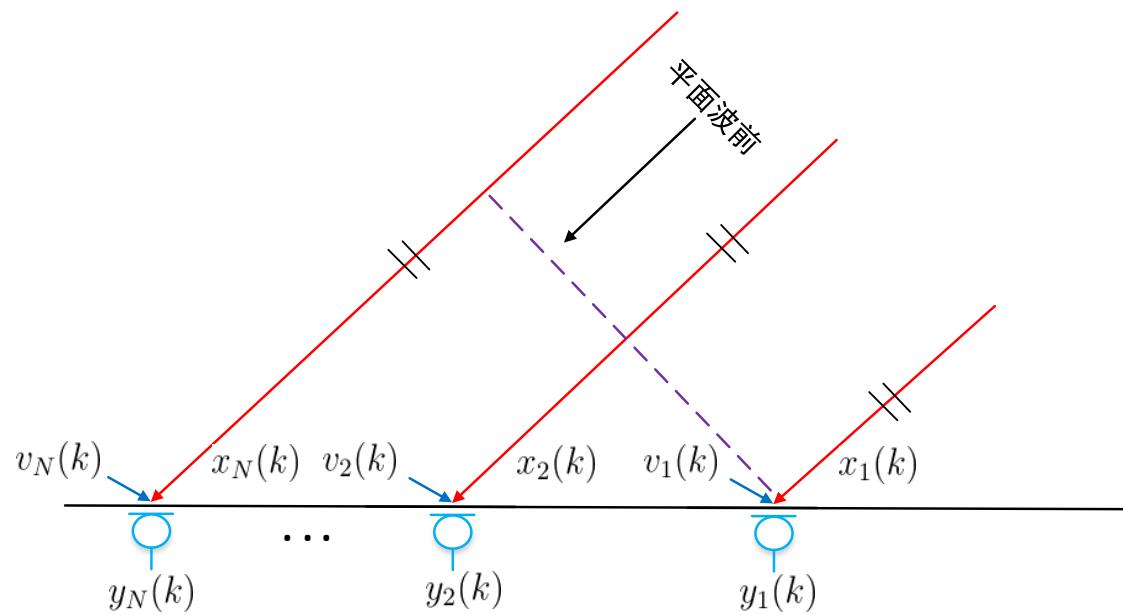
- The effect of the room can be modelled as a FIR filter.
- If we can estimate this filter, we can subtract the echo from the recorded signal before sending.



# Communication Acoustics

## Source Localization

- Sound source localization aims to determine the spatial location of a sound source in the environment based on the acoustic signals received by a set of microphones.



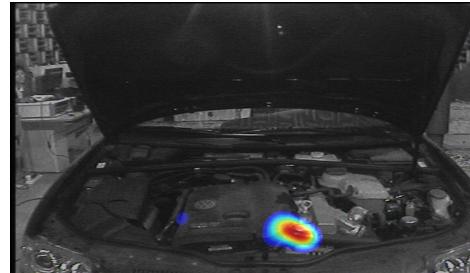
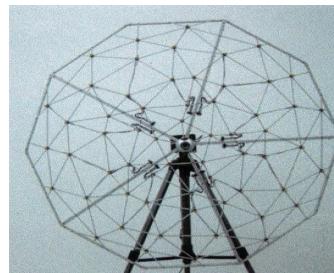
$$\begin{aligned}x_1(n) &\square s(n) \\x_2(n) &\square s(n-D) \\\vdots \\x_N(n) &\square s(n-(N-1)D) \\D &\square ?\end{aligned}$$



# Communication Acoustics

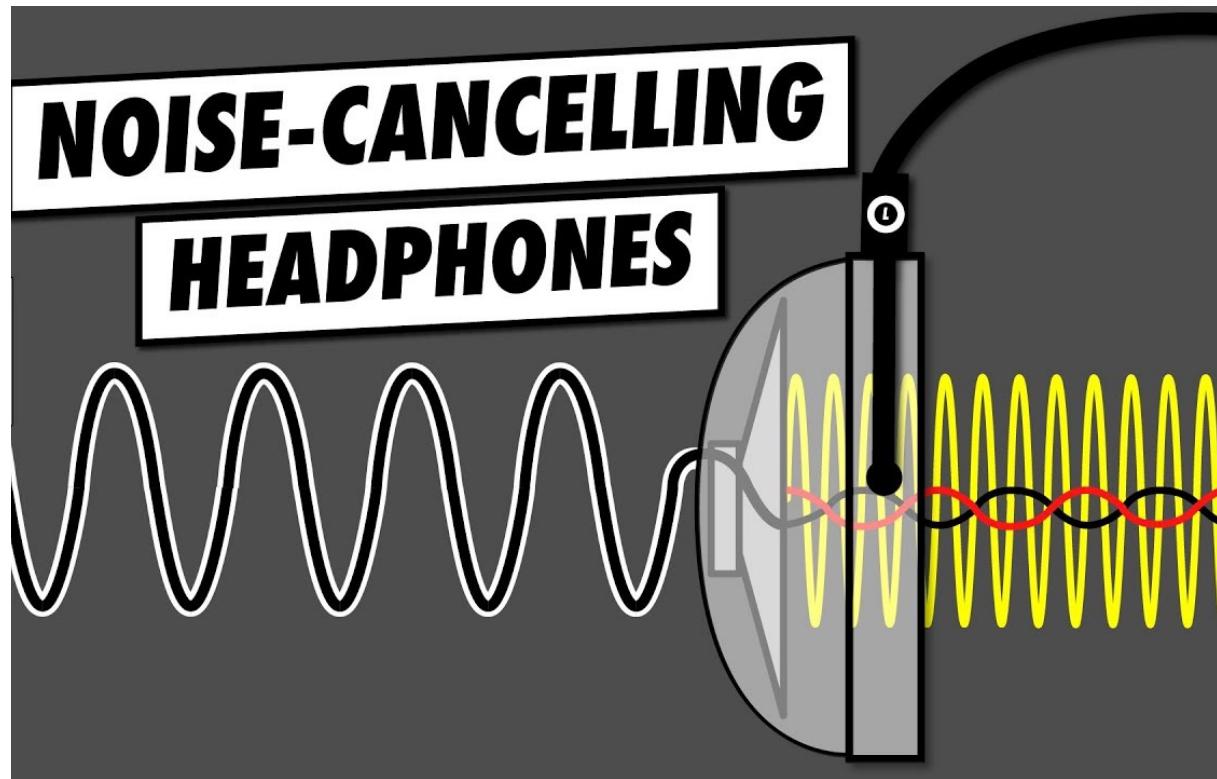
## Source Localization - Sound Camera

- Use a microphone array to measure the distribution of sound fields. It can be used to measure the location of sound emitted by an object and the state of sound radiation.



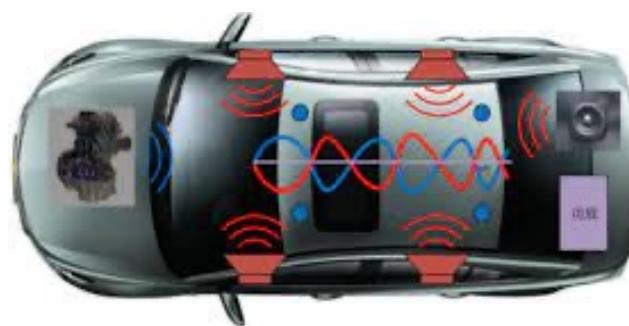
# Communication Acoustics

## Active Noise Control



# Communication Acoustics

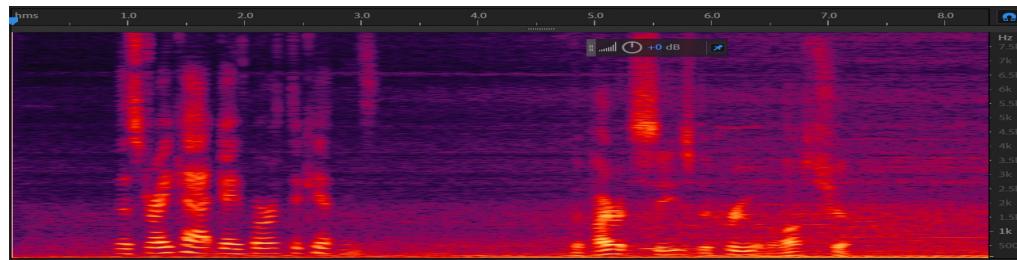
## Active Noise Control - Example



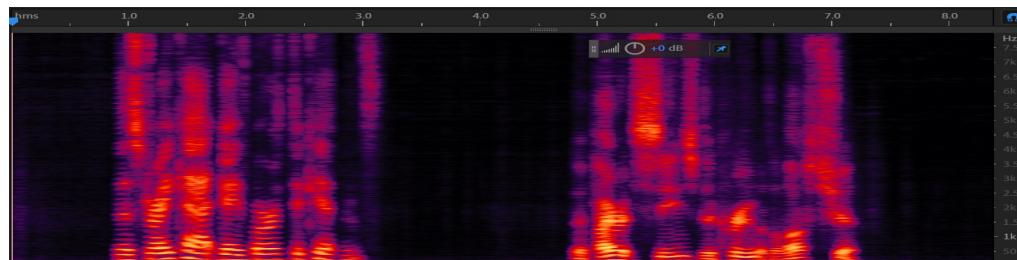
# Communication Acoustics

## Speech Enhancement

Improve the quality and intelligibility of a speech signal by removing various types of noise interference.



Raw Mic



Proposed



# **Communication Acoustics**

---

## **Speech Enhancement – Types of Noise**

- **Stationary noise**
- **Quasi stationary noise**
- **Non-stationary noise**

**Common Types :** white, pink, music, babble, bus, car, metro, office, railway, restaurant, street, traffic, workshop, airport, station



# Communication Acoustics

## Source Separation



Cocktail  
Party  
Problem



Music  
Analysis



Speech  
Recognition  
in Noisy  
Environment



Security  
Monitoring



# Communication Acoustics

## Source Separation - Example



Noisy Speech

Target Speech



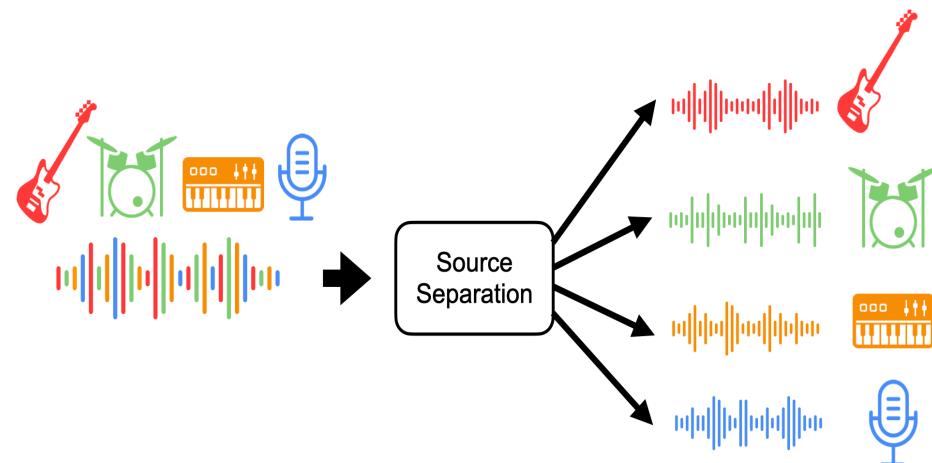
Mixture



Est1



Est2



Mixture



Vocal



Drum



Bass

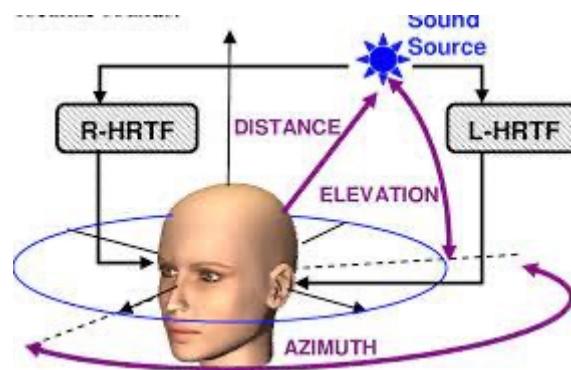
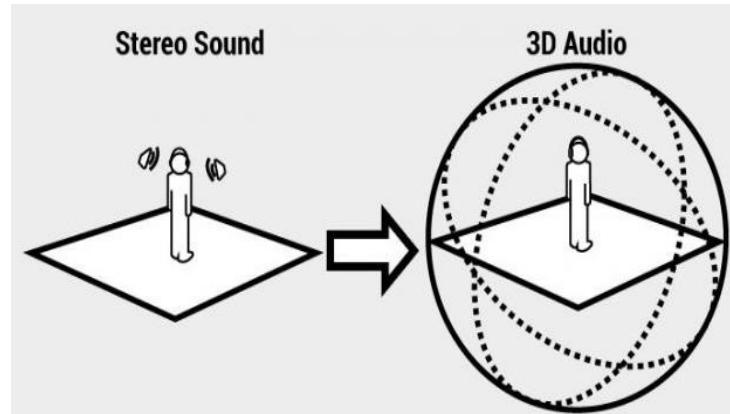


Others



# Communication Acoustics

## 3D Sound Reproduction



# Communication Acoustics

---

## Some References

- [1] R. Martin, U. Heute, C. Antweiler, Advances in Digital Speech Transmission, Wiley, 2008.
- [2] P. Naylor and D. Nikolay, eds. Speech dereverberation. Springer Science & Business Media, 2010.
- [3] B. Michael, and D. Ward, eds. Microphone arrays: signal processing techniques and applications. Springer Science & Business Media, 2001.
- [4] J. Benesty, T. Gänsler, D. R. Morgan, M. M. Sondhi, and S. L. Gay, Advances in Network and Acoustic Echo Cancellation. Springer-Verlag, Berlin, Germany, 2001.
- [5] P. Loizou. Speech enhancement: theory and practice. CRC press, 2013.



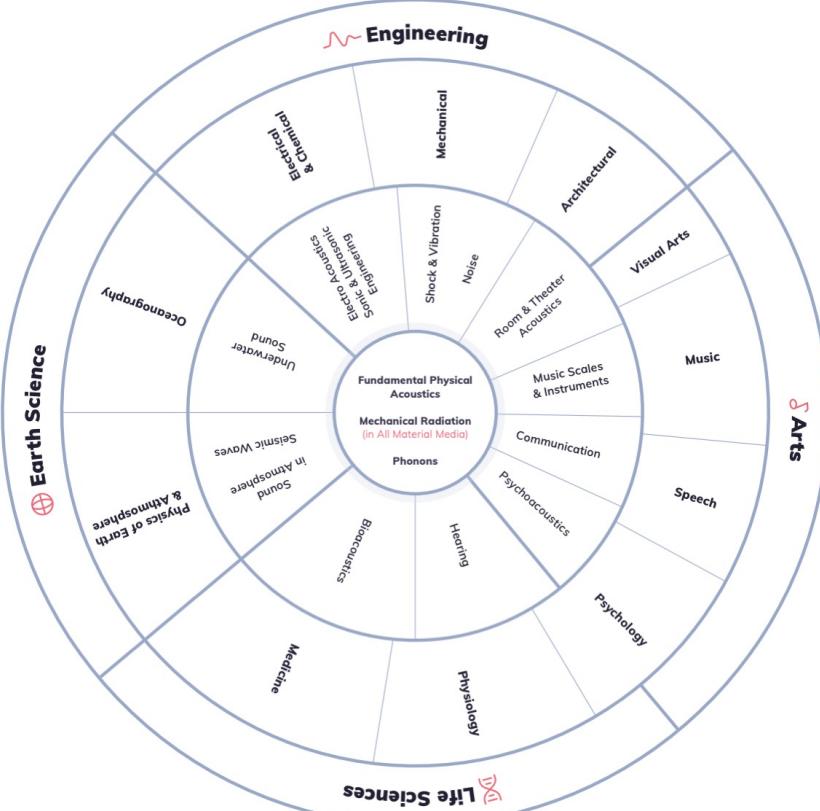
# AI for Audio Application

XJTLU | SCHOOL OF  
FILM AND  
TV ARTS



# AI for Audio Application

## Lindsay's Wheel of Acoustic



- **Audio, speech, and acoustics are collectively referred to as sound.**
- **It is the second largest application of artificial intelligence.**



# AI for Audio Application

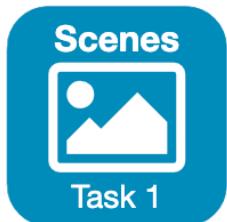
## DCASE Challenge

- Approach: “A rising tide lifts all boats”
- Before 2011: No “home” for everyday sound recognition, little data
- 2012: Plan, collect sounds, “IEEE AASP Challenge”, WASPAA slot
- DCASE 2013, Oct 2013 (at WASPAA): 24 submissions
- DCASE 2016, Sep 2016 (at EUSIPCO): 82 submissions
- DCASE 2017, Nov 2017, Munich: 200+ submissions
- DCASE 2018, Nov 2018, Surrey, UK: 650+ submissions
- DCASE 2019, Oct 2019, New York: 1000+ submissions



# AI for Audio Application

## DCASE Challenge



**Acoustic  
Scene  
Classification**



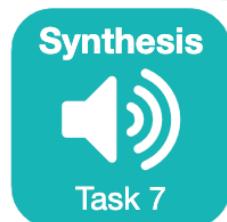
**Weakly  
Supervised  
Sound Event  
Detection**



Machine  
Condition  
Monitoring



Bioacoustics  
Event  
Detection



**Foley Sound  
Synthesis**



**Sound Event  
Localization  
and Detection**



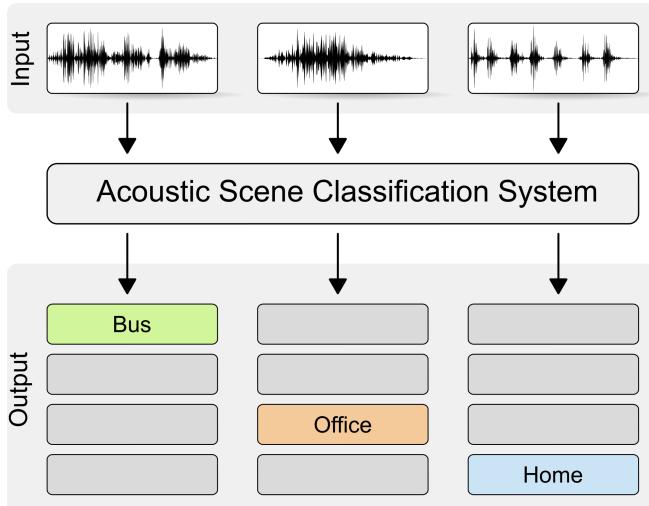
Audio  
Captioning



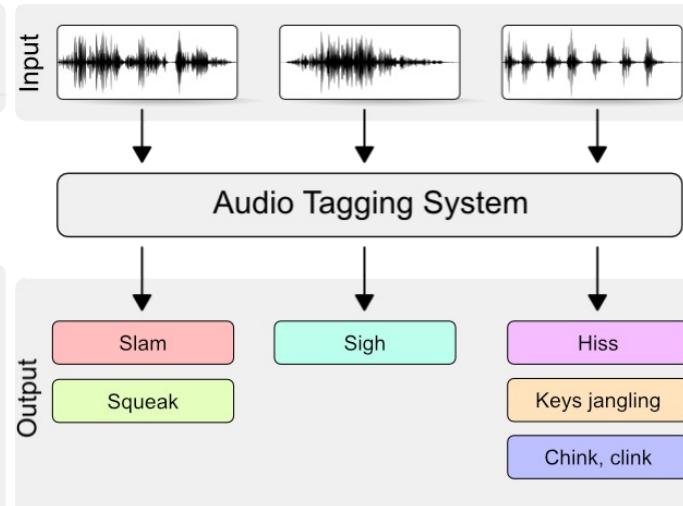
# AI for Audio Application

## Computational Analysis of Sound Scenes and Events

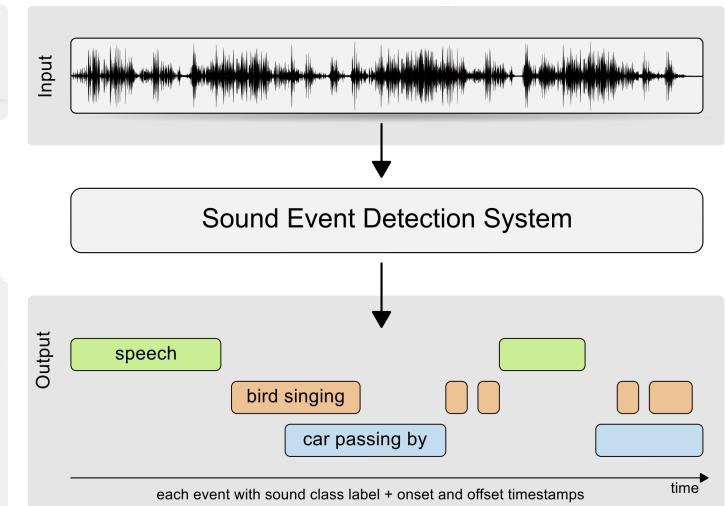
### Acoustic Scene Classification



### Audio Tagging

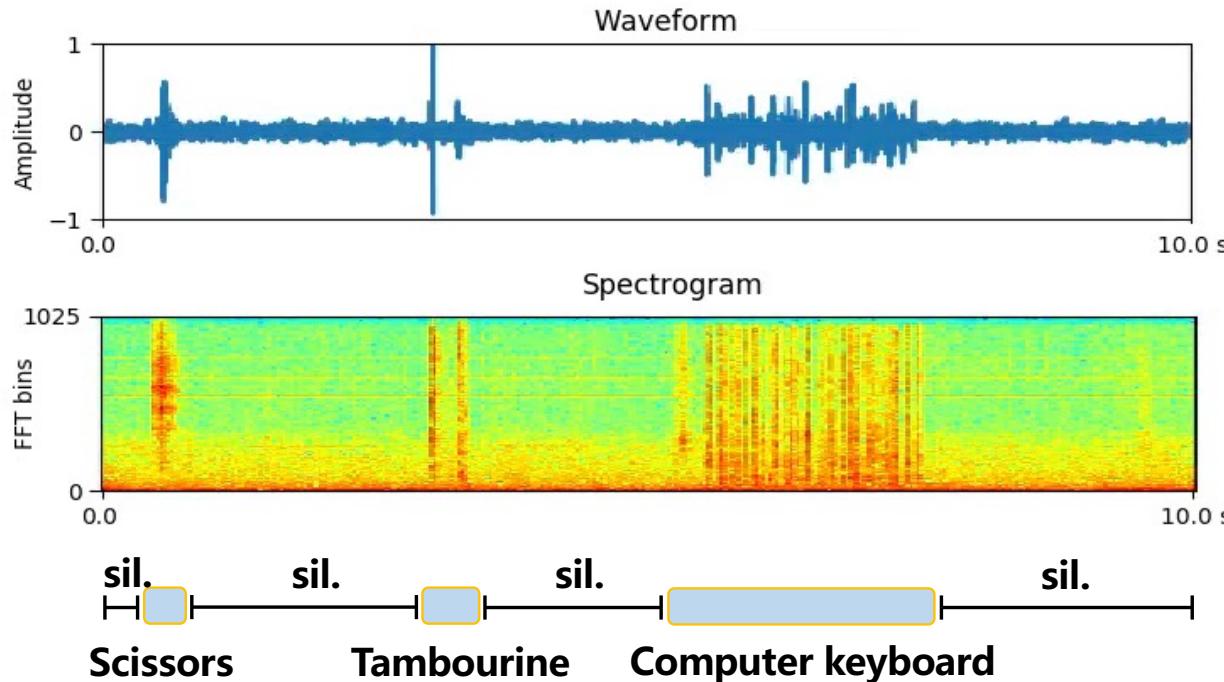


### Sound Event Detection



# AI for Audio Application

## Strong Labeled Data is Costly to Acquire



The goal of Sound Event Detection(SED):

The goal of Audio Tagging(AT):

Computer keyboard, Tambourine, Scissors

Strongly labeled data

Weakly labeled data

# AI for Audio Application

A Large Amount of Weakly Labeled Data is Available

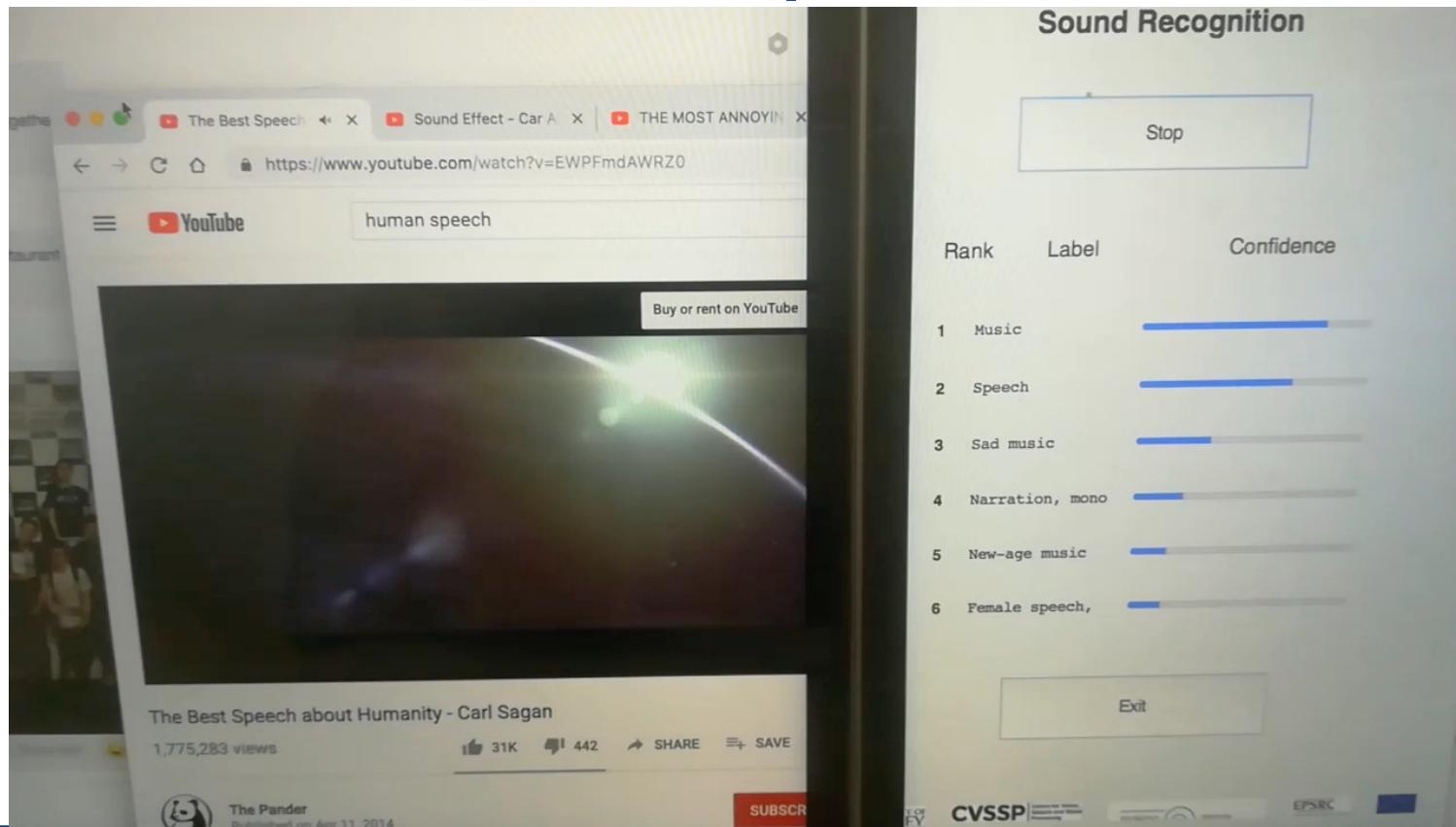


腾讯视频  
不负好时光



# AI for Audio Application

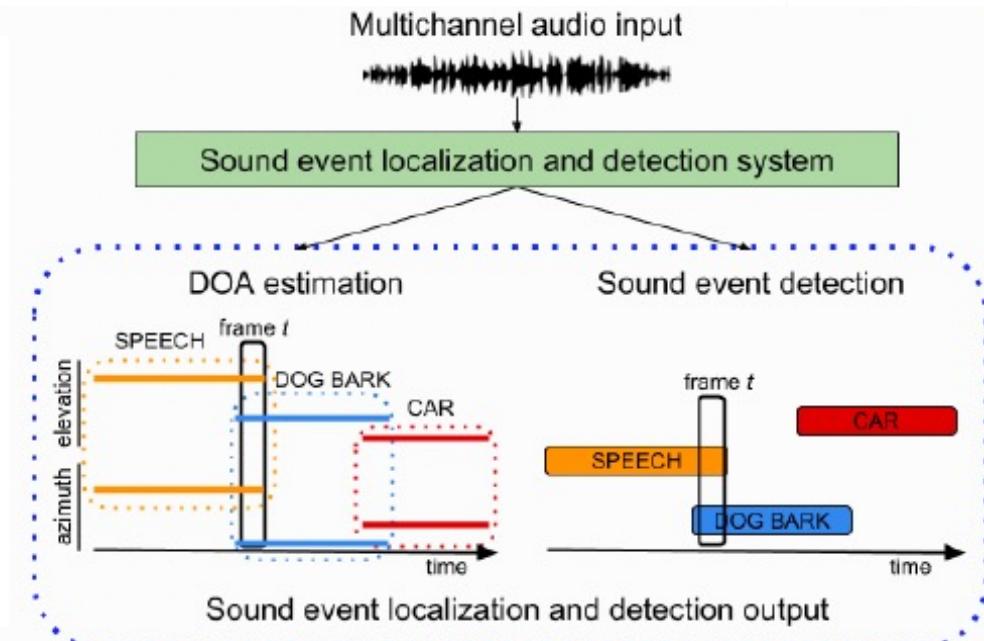
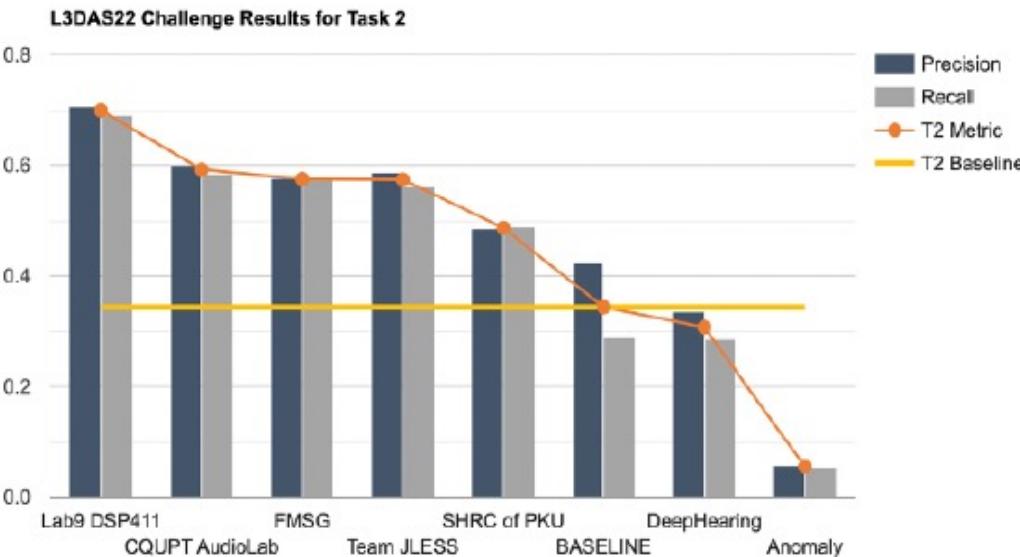
## Sound Event Detection - Example



# AI for Audio Application

## Sound Event Localization and Detection

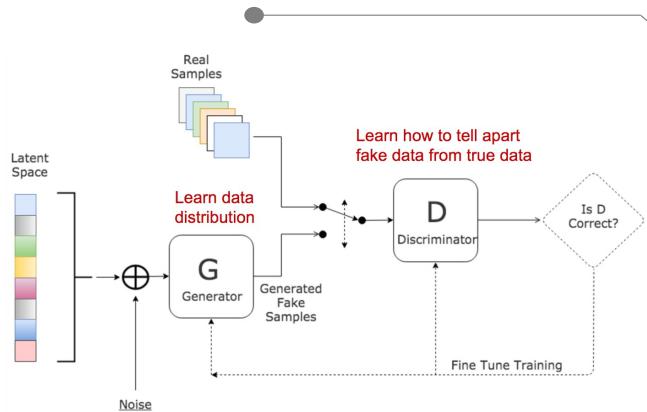
Detect the category, start and end time of ongoing acoustic events, and locates their spatial position and direction.



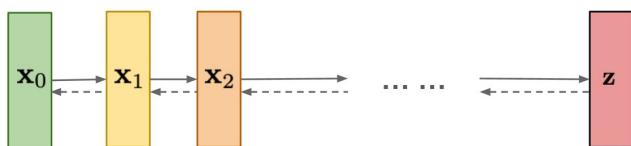
# AI for Audio Application

## Generative Models

### 对抗生成网络 (GANs)

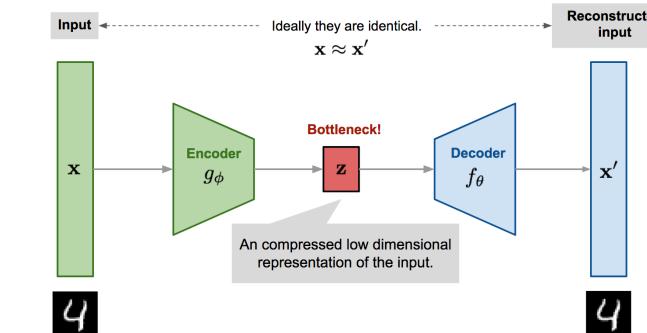


### 扩散模型 (Diffusion)



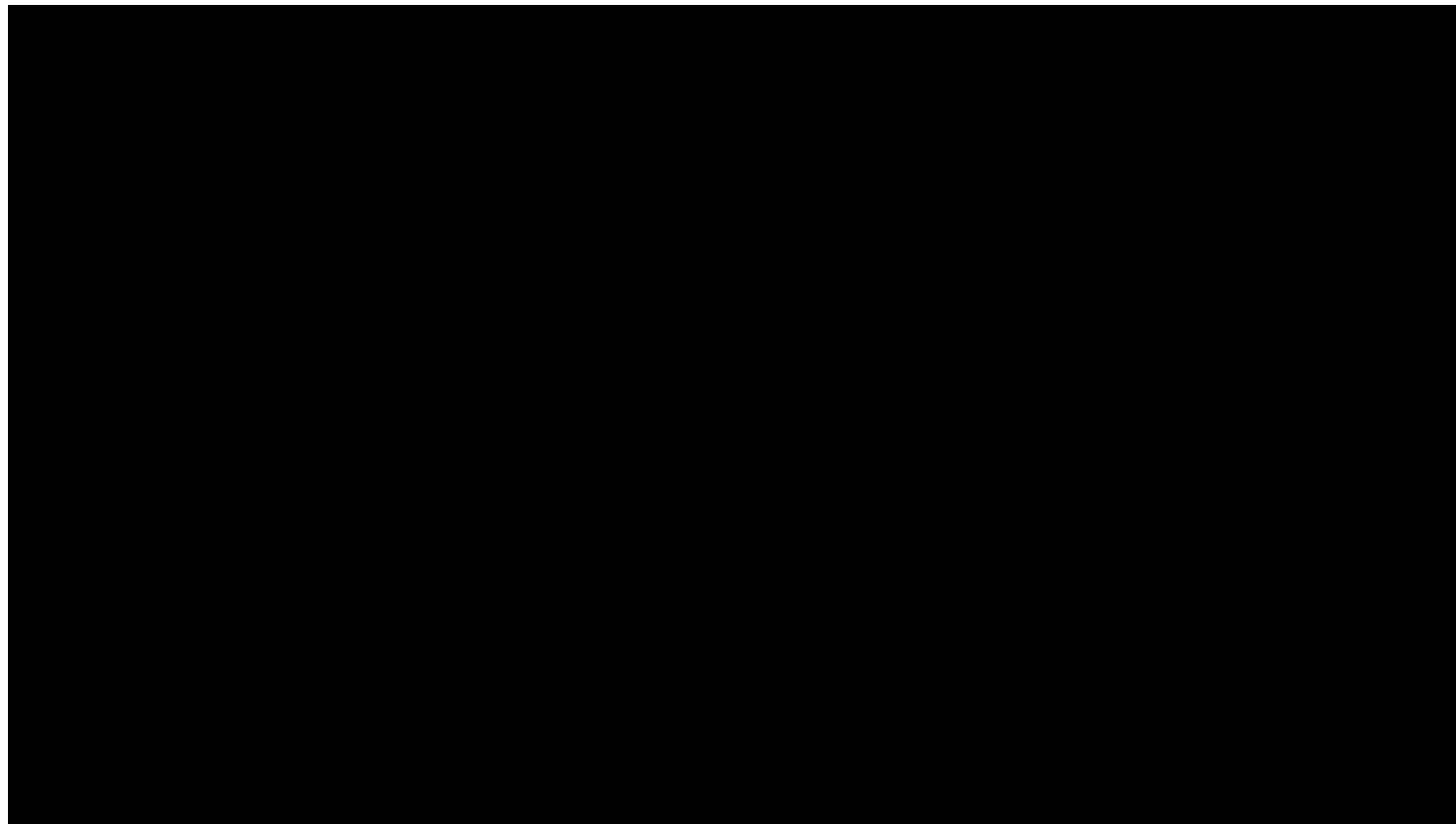
Diffusion models:  
Gradually add Gaussian  
noise and then reverse

### 自编码器与变分自编码器 (AE, VAE)



# AI for Audio Application

## Generative Models - Example





# THANK YOU



VISIT US

[WWW.XJTLU.EDU.CN](http://WWW.XJTLU.EDU.CN)



FOLLOW US

@XJTLU



Xi'an Jiaotong-Liverpool University  
西交利物浦大学

XJTLU | SCHOOL OF  
FILM AND  
TV ARTS

