Interspeech 2025 Survey Talk

Assistive Oral Communication Technology

Yu Tsao (Ph.D., Professor)
Research Center for Information Technology Innovation
Academia Sinica
yu.tsao@citi.sinica.edu.tw, https://www.citi.sinica.edu.tw/pages/yu.tsao/

August 2025

中央研究院
ACADEMIA SINICA

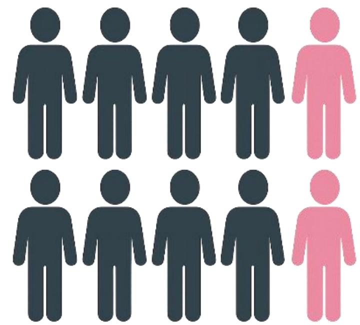
Outline

1. Research Objectives
2. Assistive Hearing Technologies
   – Diagnosis of hearing disorders
   – Speech enhancement for hearing impairment
3. Assistive Speaking Technologies
   – Diagnosis of speaking disorders
   – Speech enhancement for speaking impairment
4. Summary and Future Directions

Research Objectives

Issues of Hearing Loss

Present



> 65y 40%

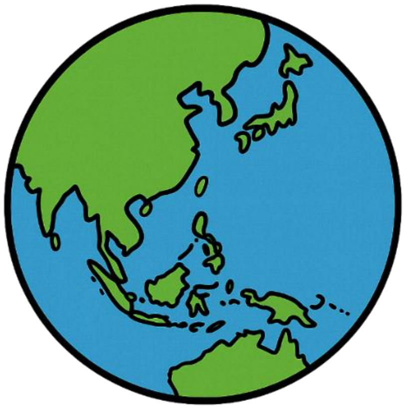
Isolation

Dementia

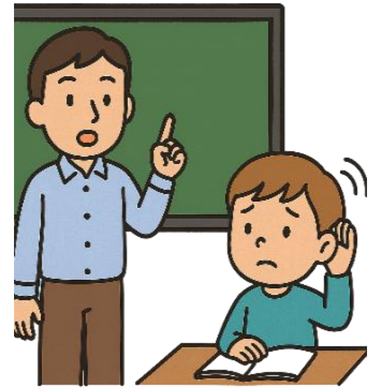
Stigma

Learning difficulty

Issues of Speaking Disorder



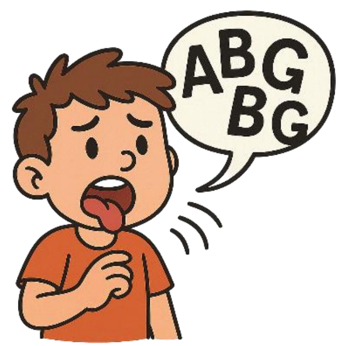
> 170 million



Hearing Loss

Stuttering

1%



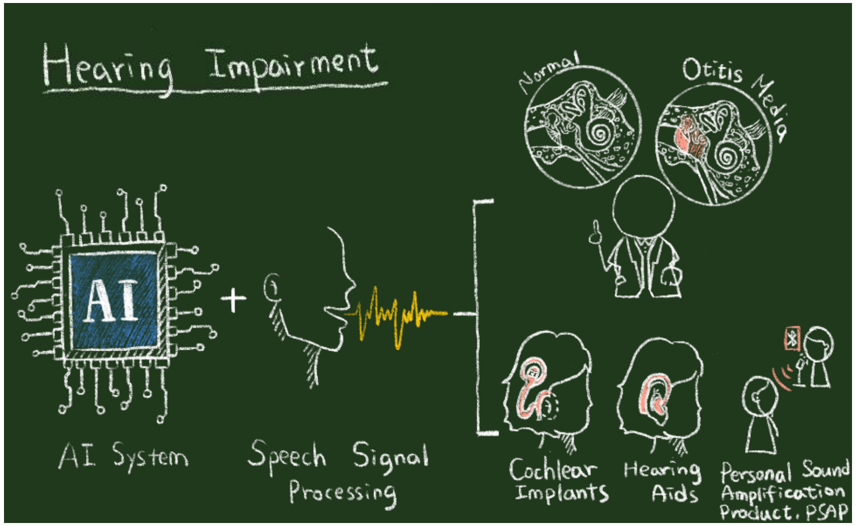
Dysarthria

3-7 %

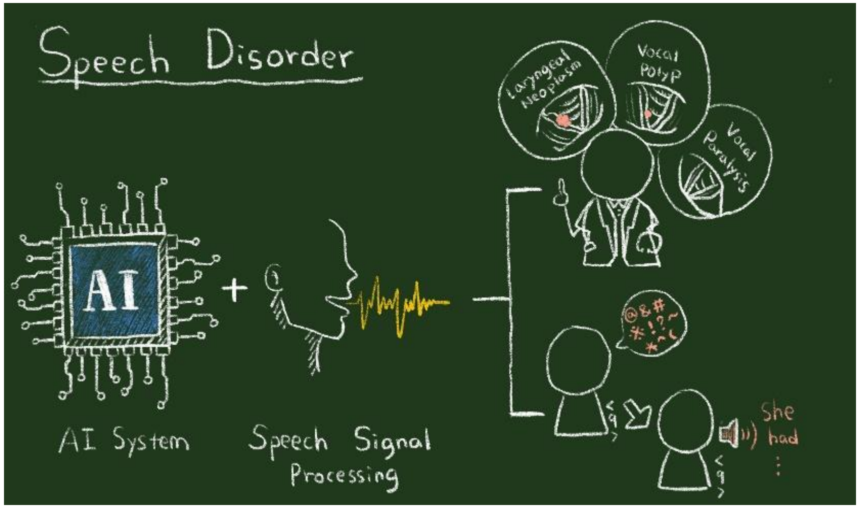
Surgery

> 300,000 users

Assistive Oral Communication Technologies



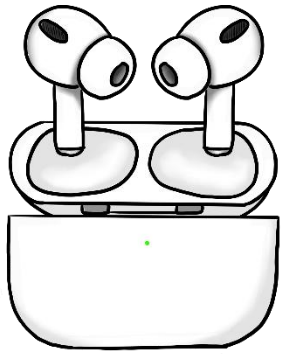
- **Assistive Hearing Technologies**
  - ➢ Diagnosis
    - ✓ Automatic otitis media with effusion (OME) detection
  - ➢ Speech Enhancement
    - ✓ Speech processing for HA/CI
    - ✓ Multimodal processing for HA/CI
    - ✓ Speech assessment for hearing loss



- **Assistive Speaking Technologies**
  - ➢ Diagnosis
    - ✓ Pathological voice detection
    - ✓ Pathological voice assessment
  - ➢ Speech Enhancement
    - ✓ Dysarthric speech
    - ✓ Oral surgical speech
    - ✓ Electrolaryngeal speech

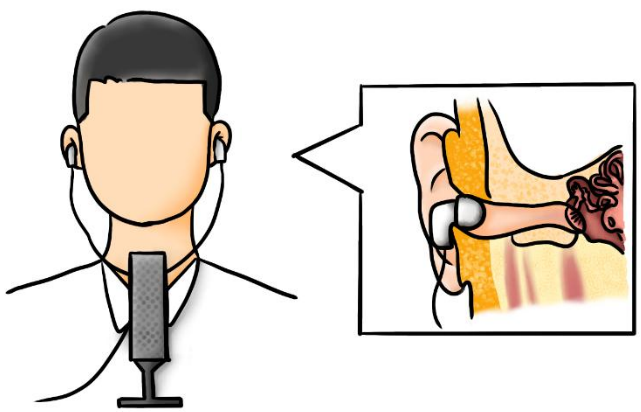
Automatic Detection of Otitis Media with Effusion (OME)

In-Ear Microphone

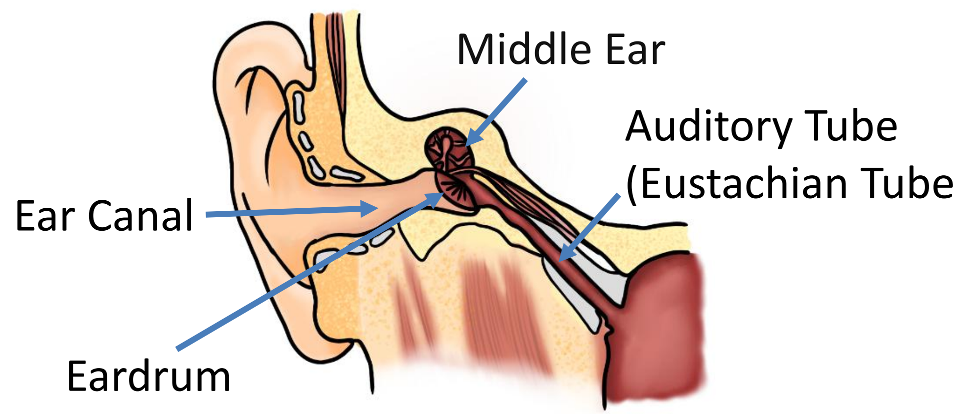


Active Noise Cancelling (ANC) Headset

ANC Headset

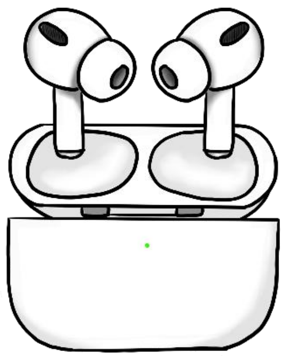


In-ear Microphone



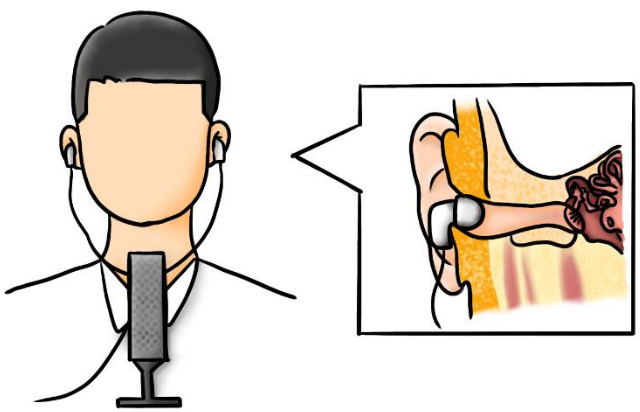
Anatomy Chart

In-Ear Microphone

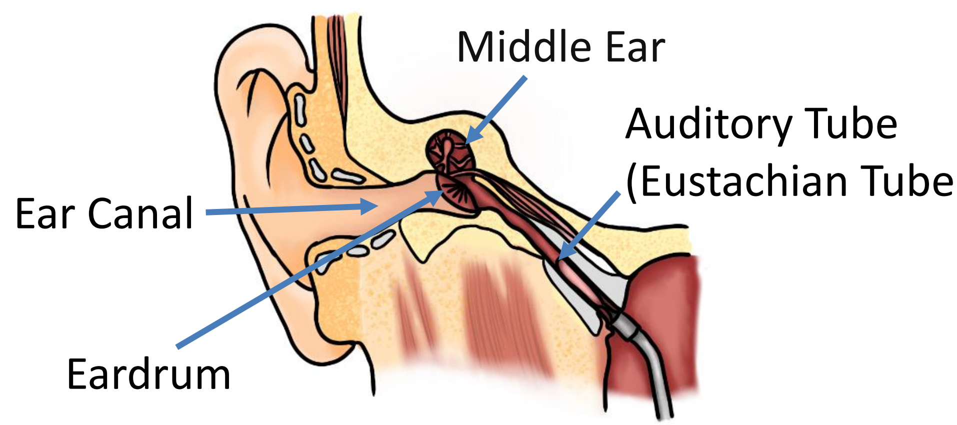


Active Noise Cancelling (ANC) Headset

ANC Headset

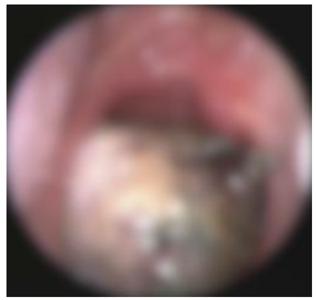


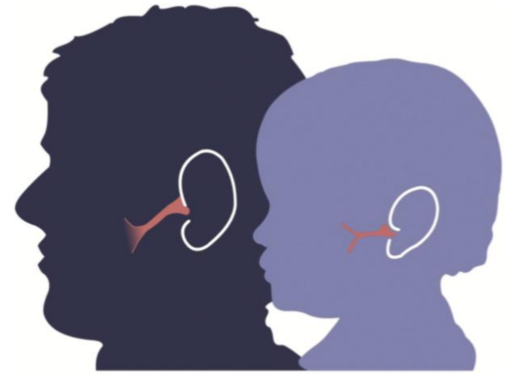
In-ear Microphone



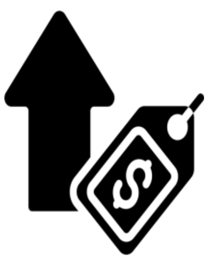
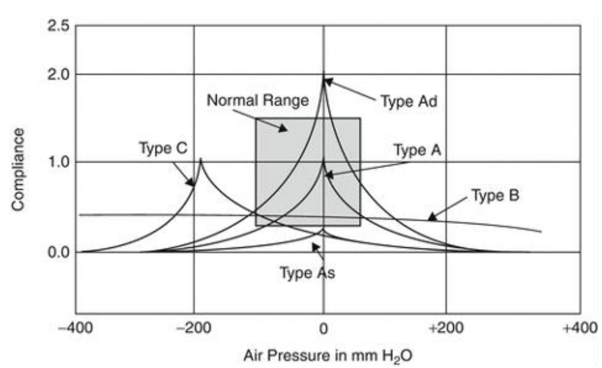
Anatomy Chart

Otitis Media with Effusion (OME) Detection



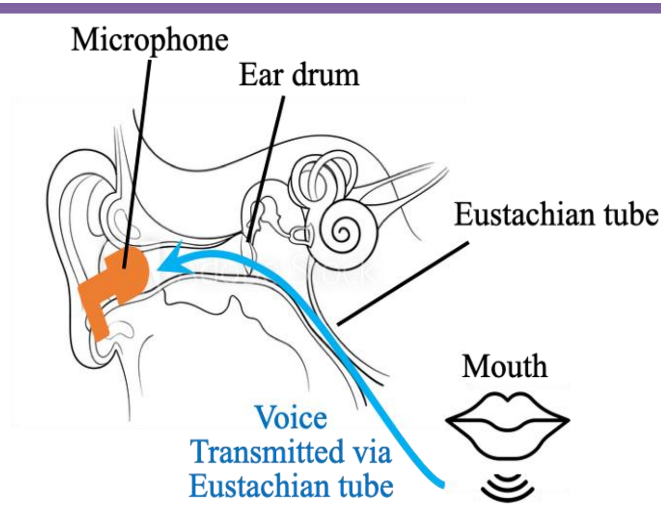


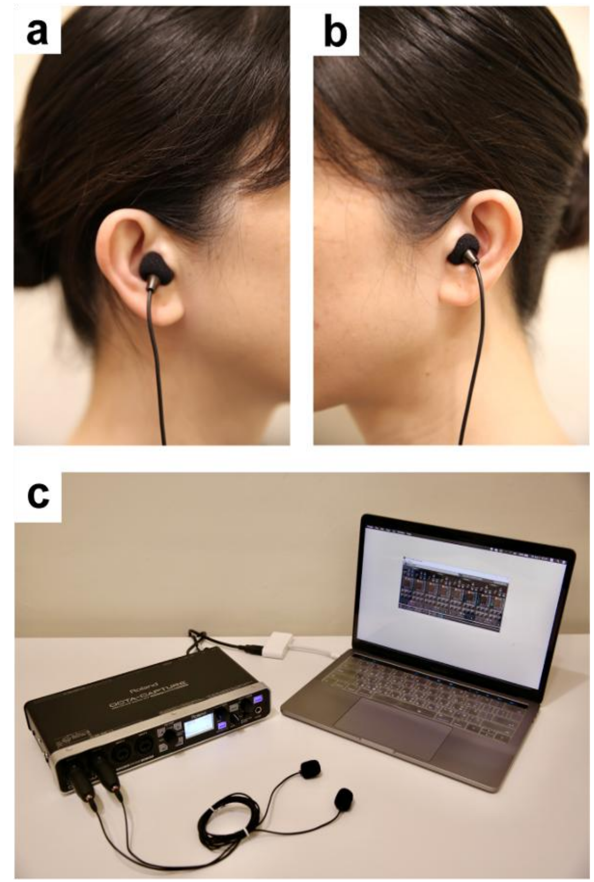



- Physical examination
  - Otoscope
  - Telescope
  - Tympanometry
- Challenges
  - Specialists' interpretation
  - High-cost: $4 billion
  - Inability to monitor progress
- At least 25% of OME episodes persist for > 3 months.

Otitis Media with Effusion (OME) Detection

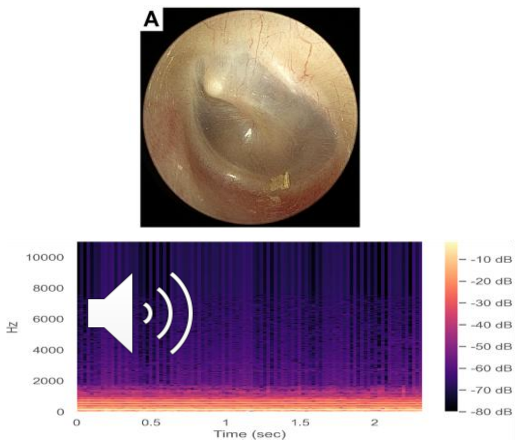
- The proposed approach aims at
  - Easy-to-use
  - Low-cost
  - Automatic



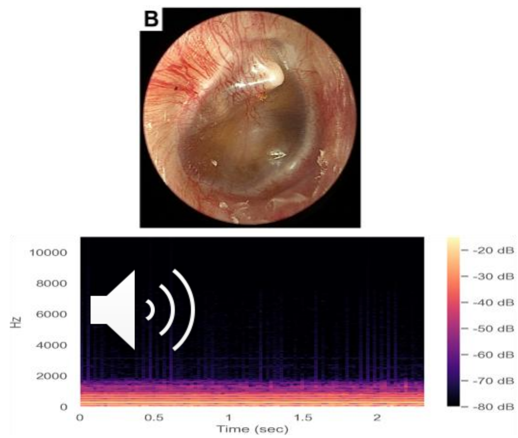


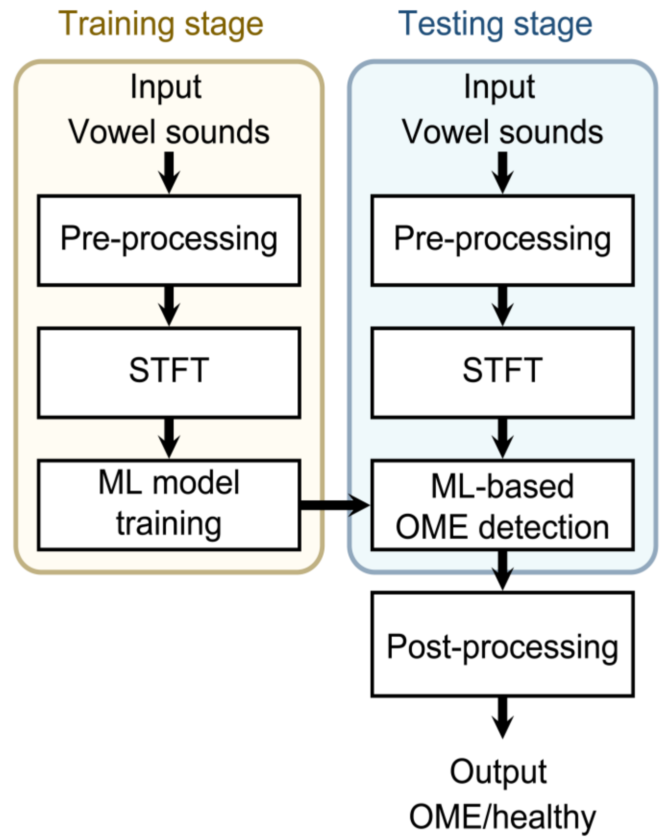
Otitis Media with Effusion (OME) Detection

Healthy ear



Ear with OME





Detection Technique	Performance Measure (%)				
	ACC	SEN	PRE	SPE	β^a
SVM	67.74	78.38	70.73	52.73	13
GNB	62.90	72.97	67.50	48.08	32
AdaBoost	74.19	75.68	80.00	72.67	10
RF	62.90	94.59	62.50	16.32	13
CNN	**80.65**	66.67	73.68	87.80	23

CNN has the highest accuracy

[1] Detection of Otitis Media With Effusion Using In-Ear Microphones and Machine Learning, IEEE Sensors Journal, vol. 23, no. 22, pp. 28411-28420, 15 Nov.15, 2023.

Related Works

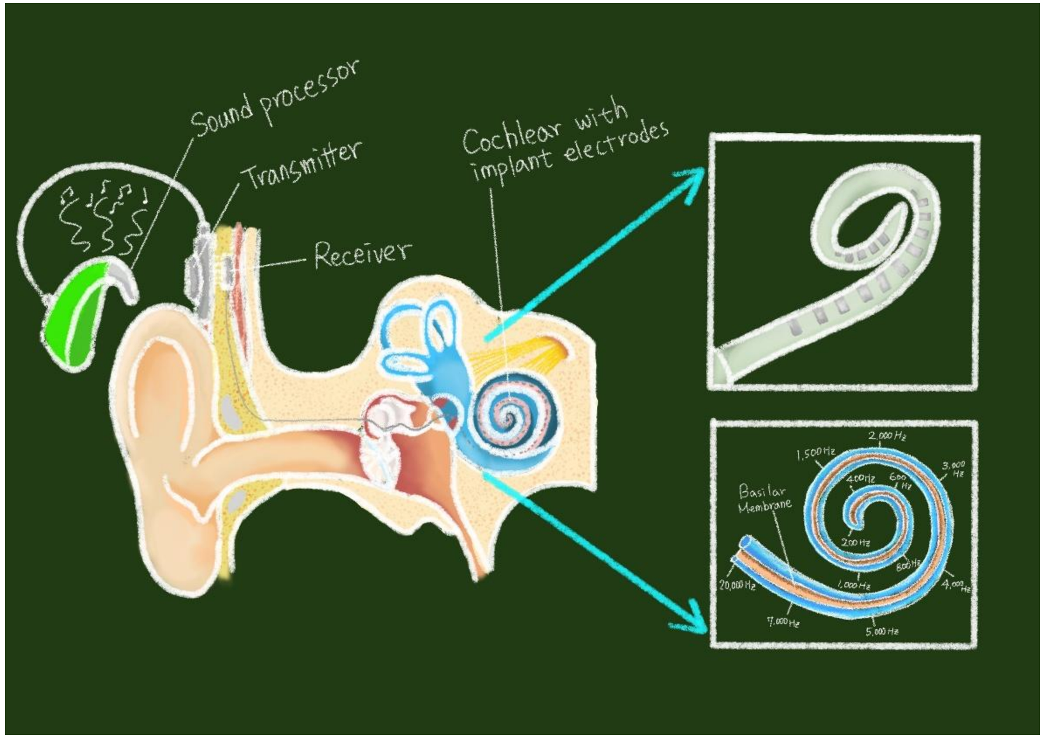
No	Title	Diagnosis	Data
1	A Hybrid Deep Learning Approach to Identify Preventable Childhood Hearing Loss	Middle Ear	Tympanometry
2	Enhanced Diagnosis of Otitis Media with Effusion using Machine Learning	Middle Ear	Acoustic Signals
3	Detecting Middle Ear Fluid using Smartphones	Middle Ear	Acoustic Signals
4	Inertial Measurement Unit-Based Romberg Test for Assessing Adults with Vestibular Hypofunction	Vestibular System	Inertial Data
5	Machine Learning Models Can Predict Tinnitus and Noise-Induced Hearing Loss	Tinnitus	Audiometry + Clinical Data
6	Deafness Gene Screening Based on a Multilevel Cascaded BPNN Model	Deafness	Genetic Data
7	A Machine Learning Screening Model for Identifying the Risk of High-Frequency Hearing Impairment in a General Population	Hearing Loss	Biomarker + Clinical Data
8	A Noise-Induced Hearing Loss Prediction Model Based on Asymmetric Convolution for Workers Exposed to Complex Industrial Noise	Hearing Loss	Speech + Clinical Data
9	Analysis of Related Factors Between Sudden Sensorineural Hearing Loss and Serum Indices Based on Artificial Intelligence and Big Data	Hearing Loss	Biomarker
10	Applying Machine Learning Algorithms to Explore the Impact of Combined Noise and Dust on Hearing Loss in Occupationally Exposed Populations	Hearing Loss	Demographic Clinical Data

Related Works

No	Title	Diagnosis	Data
11	Automated Hearing Loss Type Classification Based on Pure Tone Audiometry Data	Hearing Loss	Audiometry Data
12	Automatic Prediction of Conductive Hearing Loss Using Video Pneumatic Otoscopy and Deep Learning Algorithm	Hearing Loss	Video
13	Automatic Recognition of Auditory Brainstem Response Waveforms Using a Deep Learning-Based Framework	Hearing Loss	Brain Signals
14	Feature-Based Audiogram Value Estimator (FAVE): Estimating Numerical Thresholds from Scanned Images of Handwritten Audiograms	Hearing Loss	Image
15	Heavy Metal Biomarkers and Their Impact on Hearing Loss Risk: A Machine Learning Framework Analysis	Hearing Loss	Biomarker
16	Hybrid Statistical and Machine-Learning Approach to Hearing-Loss Identification Based on an Oversampling Technique	Hearing Loss	Demographic Clinical Data
17	Comparison Between Logistic Regression and Machine Learning Algorithms on Prediction of Noise-Induced Hearing Loss and Investigation of SNP Loci	Hearing Loss	Demographic Clinical Data
18	Machine Learning Prediction of Objective Hearing Loss With Demographics, Clinical Factors, and Subjective Hearing Status.	Hearing Loss	Demographic Clinical Subjective Data
19	Maximising the Ability of Stimulus-Frequency Otoacoustic Emissions to Predict Hearing Status and Thresholds Using Machine-Learning Models	Hearing Loss	Otoacoustic Emissions
20	Estimating Hearing Thresholds From Stimulus-Frequency Otoacoustic Emissions	Hearing Loss	Otoacoustic Emissions

Speech Enhancement for Assistive Hearing Devices

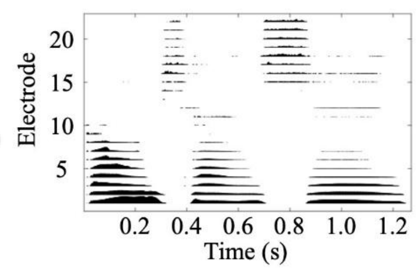
Cochlear Implant (Artificial Ear)



Input Speech (X)

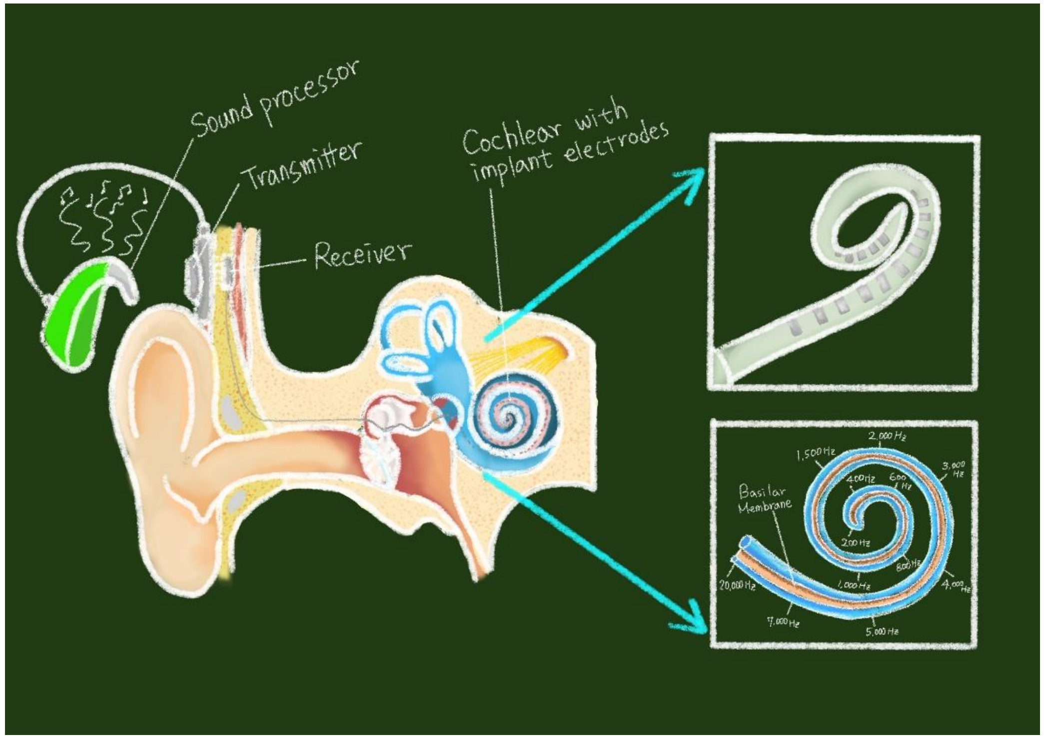
Speech Enhancement | Acoustic to Electric Conversion

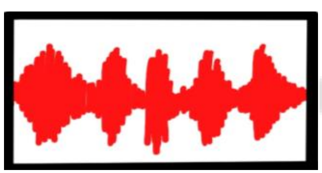
Multi-stages of Signal Processing Units



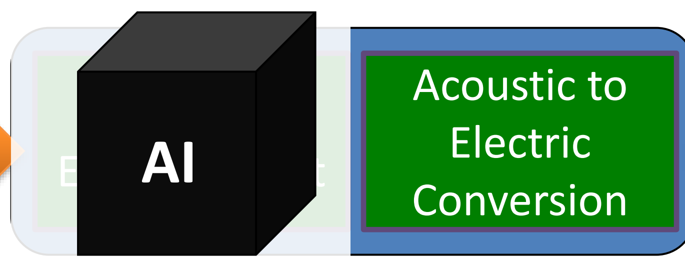
Electric Signals (Ỹ)

Cochlear Implant (Artificial Ear)

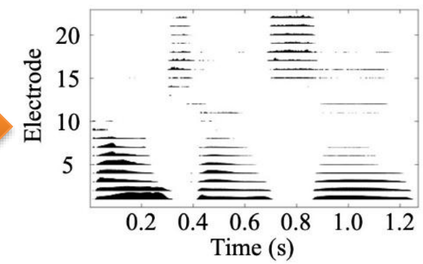




Input Speech (X)

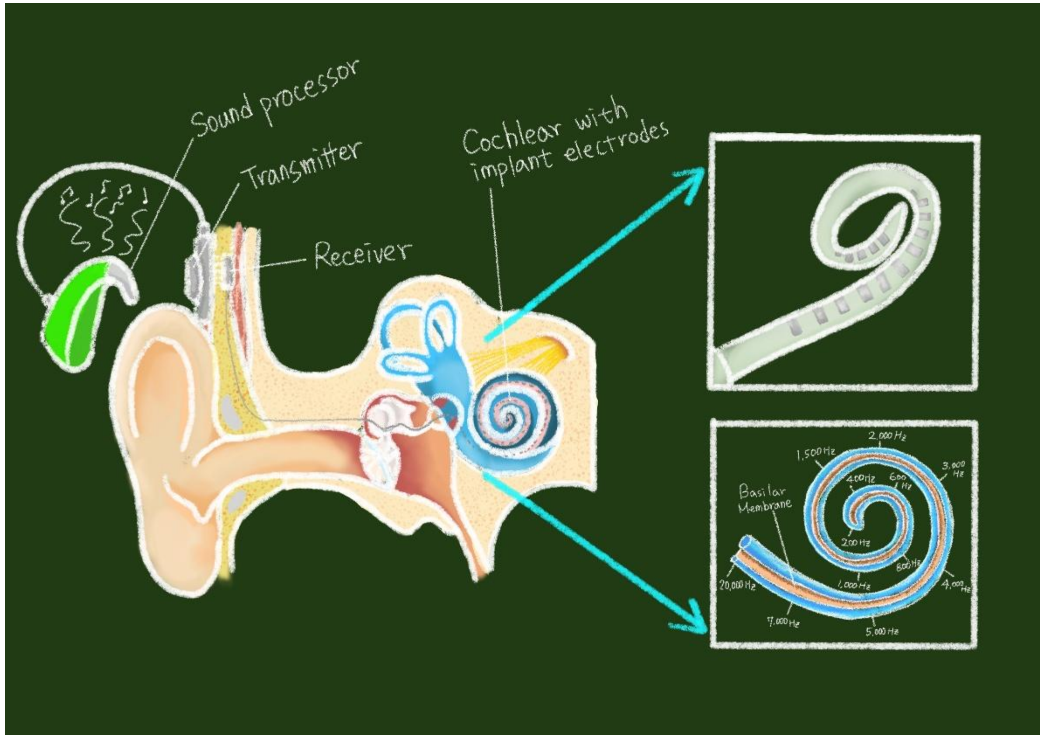


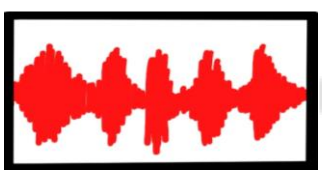
Multi-stages of Signal Processing Units



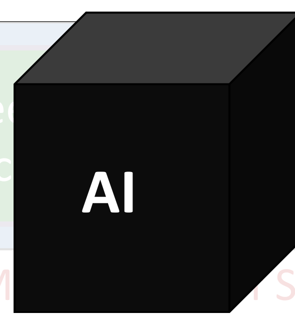
Electric Signals ($\tilde{Y}$)

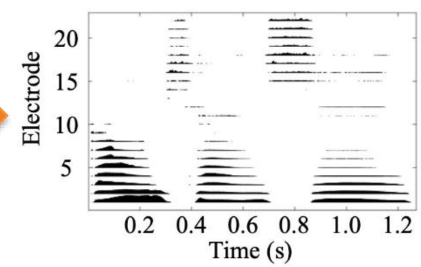
Cochlear Implant (Artificial Ear)





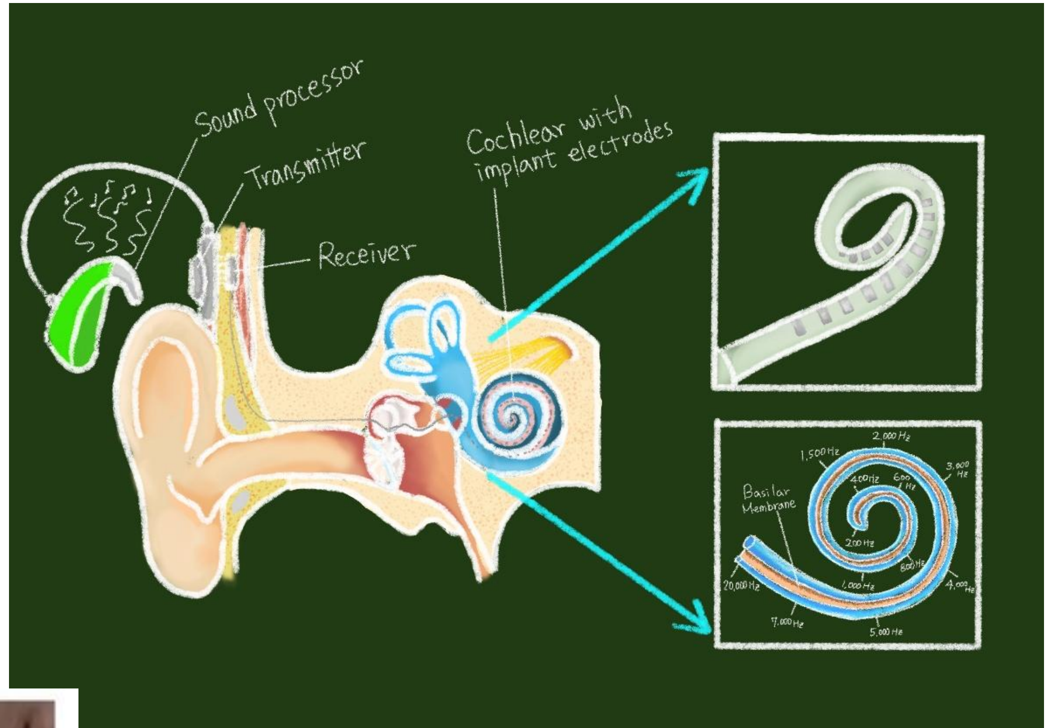
Input Speech (X)





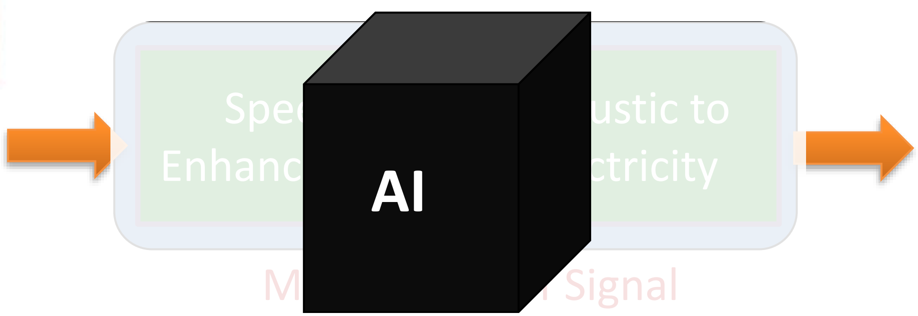
Electric Signals (Ỹ)

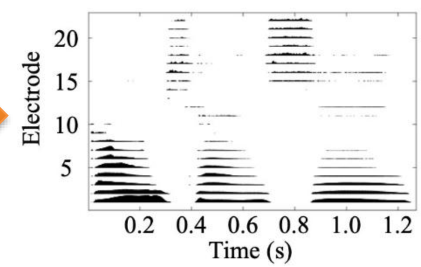
Multimodal Learning for Cochlear Implant



Visual Data (V)

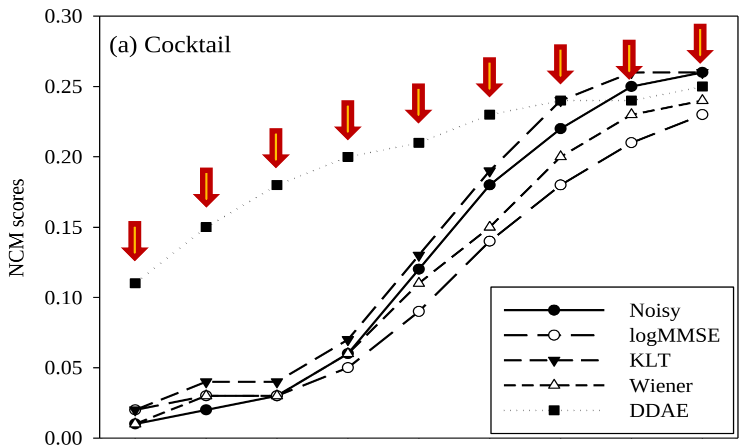
Input Speech (X)

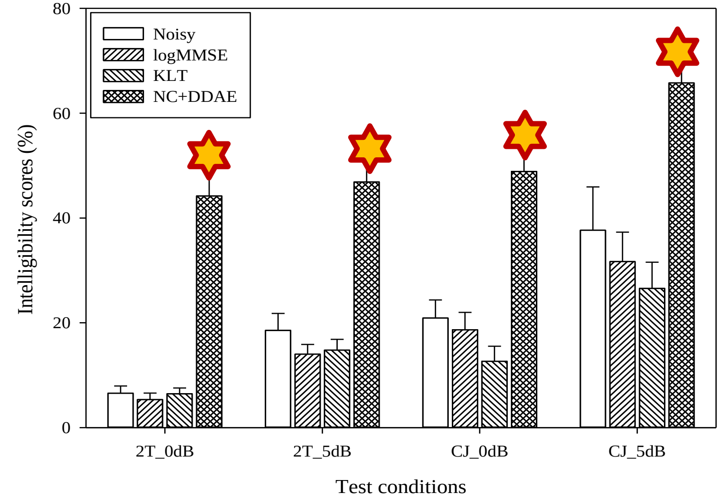


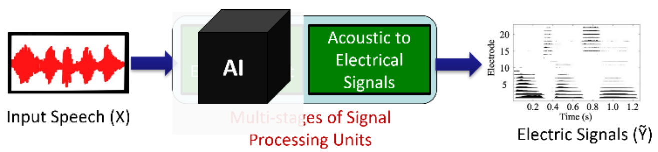


Electric Signals ($\tilde{Y}$)

Experimental Results (Cochlear Implant)



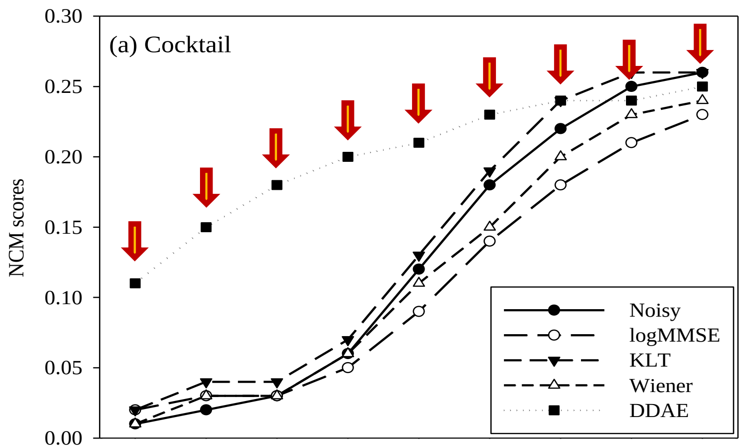


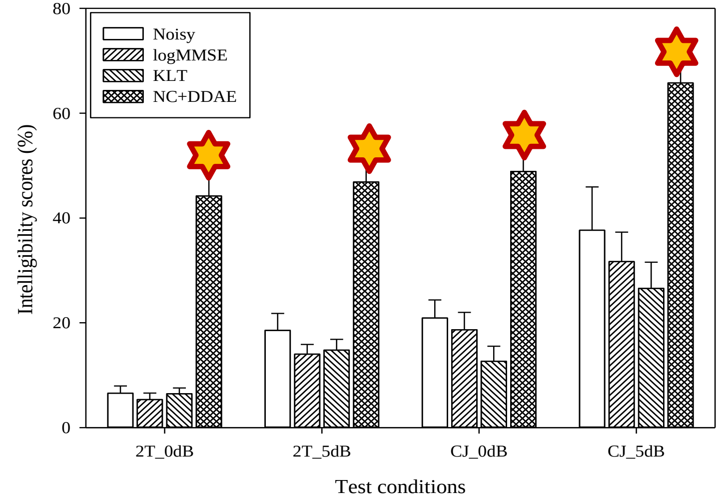


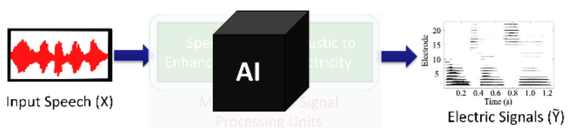
(1) Neural speech enhancement significantly **improves CI performance**.

[2] A Deep Denoising Autoencoder Approach to Improving the Intelligibility of Vocoded Speech in Cochlear Implant Simulation, IEEE TBME, vol. 64 (7), pp. 1568-1578, July 2017.

Experimental Results (Cochlear Implant)





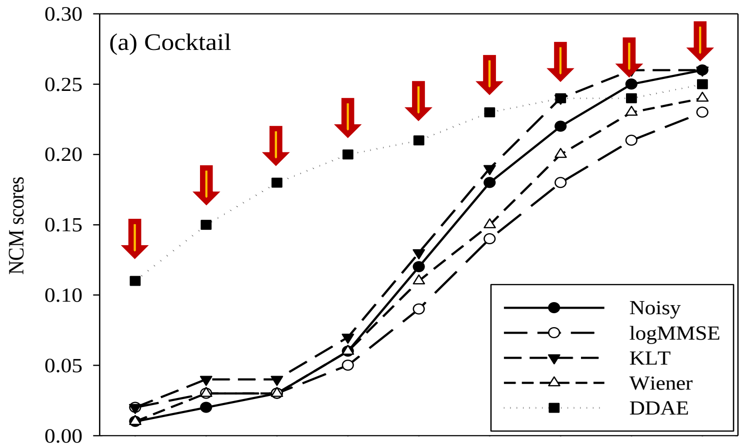


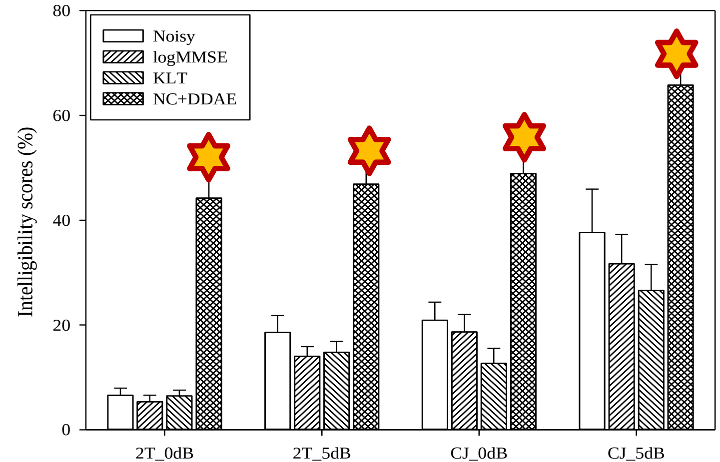
Method	Configuration	STOI↑	ESTOI↑	NCM↑
ECS	-	0.4870	0.2073	0.3258
ASE-ECS	Pre-trained	0.5943	0.3557	0.4914
AVSE-ECS	Pre-trained	0.6141	0.3672	0.5067
AVSE-ECS	Joint training	**0.6305**	**0.3899**	**0.5211**

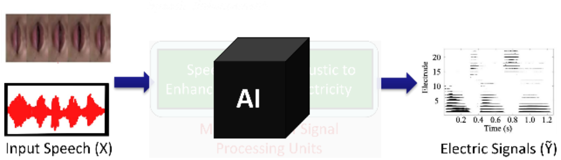
(1) Neural speech enhancement significantly **improves CI performance**.
(2) **Jointly optimizing** speech enhancement with CI systems leads to performance gains.

[3] ElectrodeNet–A Deep-learning-based Sound Coding Strategy for Cochlear Implants," IEEE Transactions on Cognitive and Developmental Systems, vol. 16 (1), pp. 346-357, 2024.

Experimental Results (Cochlear Implant)





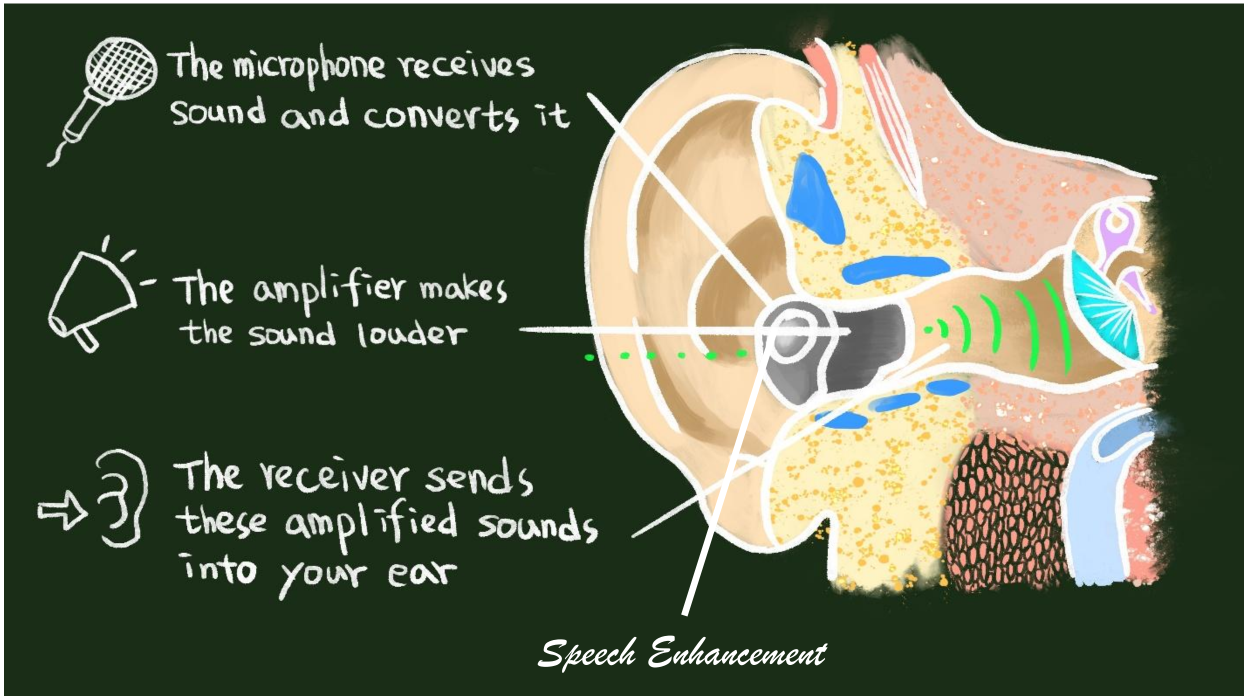


Method	Configuration	STOI↑	ESTOI↑	NCM↑
ECS	-	0.4870	0.2073	0.3258
ASE-ECS	Pre-trained	0.5943	0.3557	0.4914
AVSE-ECS	Pre-trained	0.6141	0.3672	0.5067
AVSE-ECS	Joint training	**0.6305**	**0.3899**	**0.5211**

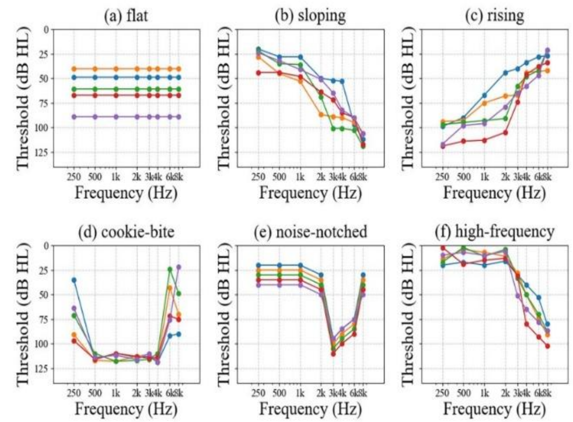
(1) Neural speech enhancement significantly **improves CI performance**.
(2) **Jointly optimizing** speech enhancement with CI systems leads to performance gains.
(3) Incorporating **visual information** yields even **greater** performance improvements.

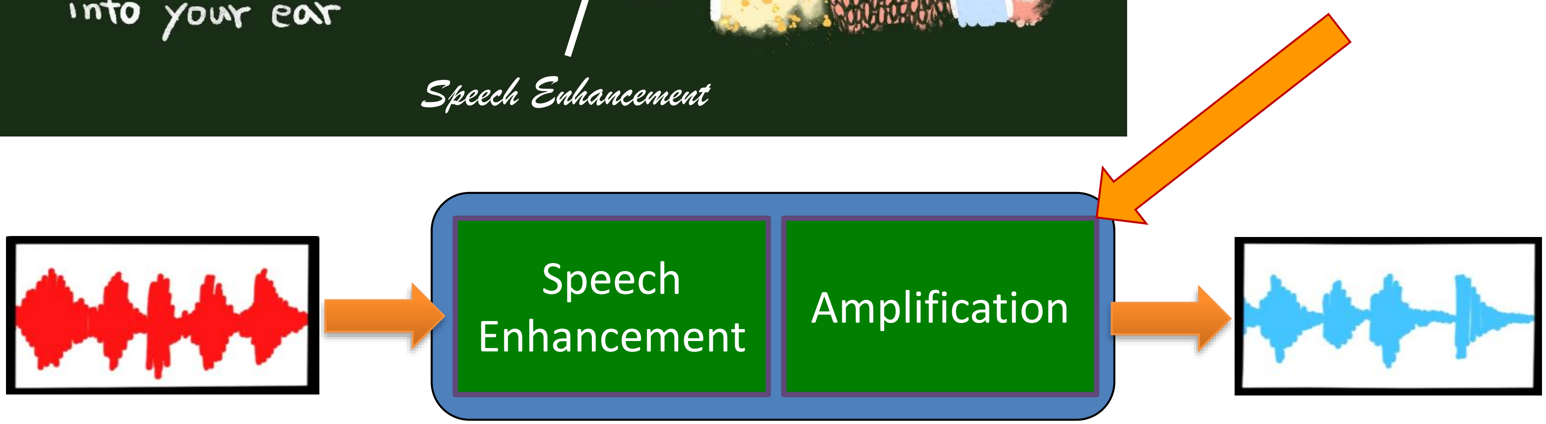
[4] End-to-End Audio-Visual Learning for Cochlear Implant Sound Coding in Noisy Environments, under review.

Hearing Aids



Audiogram (Z)



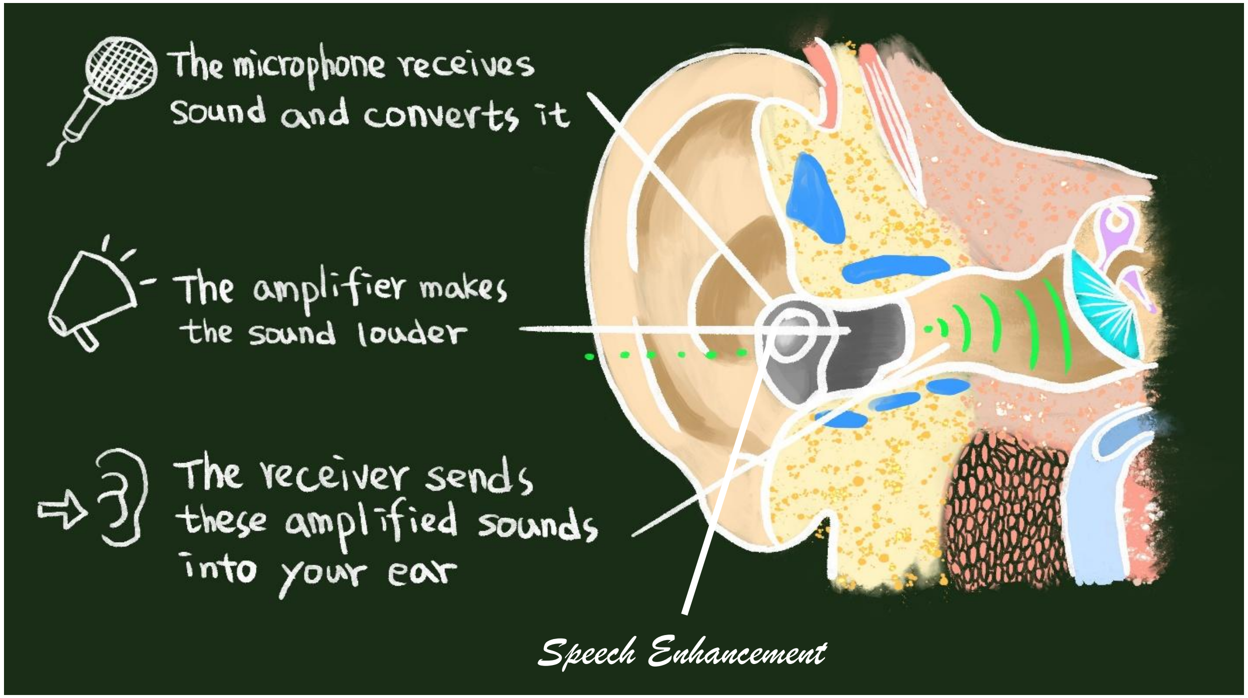


Input Speech (X)

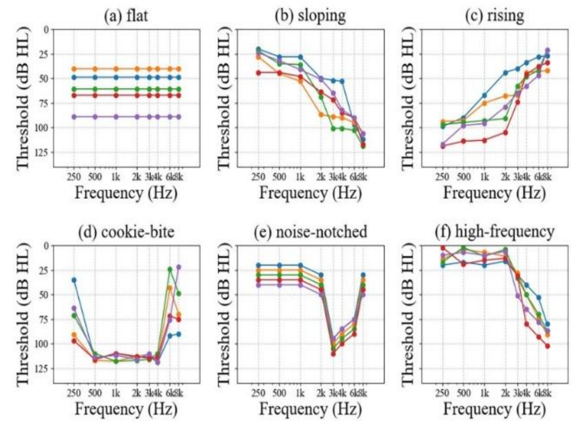
Multi-stages of Signal Processing Units

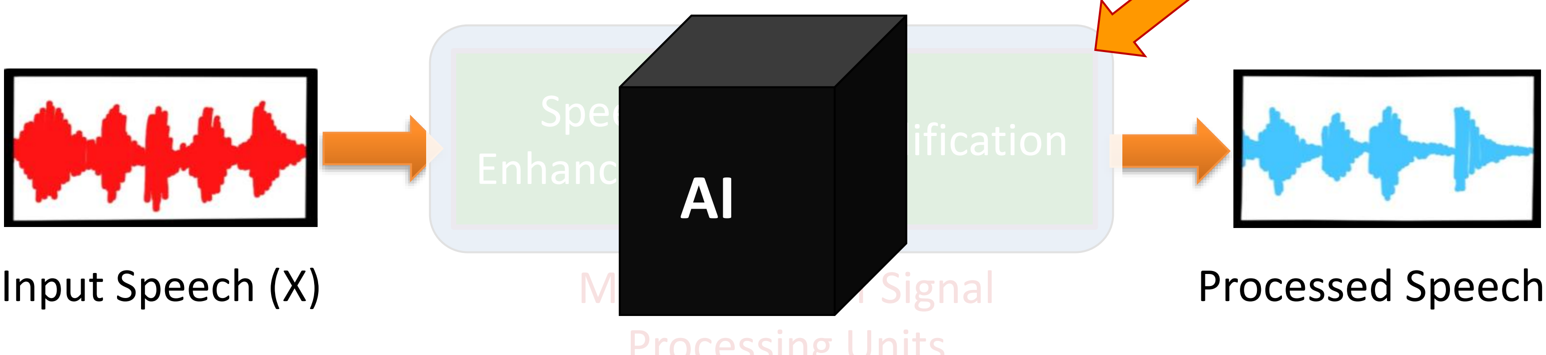
Processed Speech ($\tilde{Y}$)

Hearing Aids



Audiogram (Z)



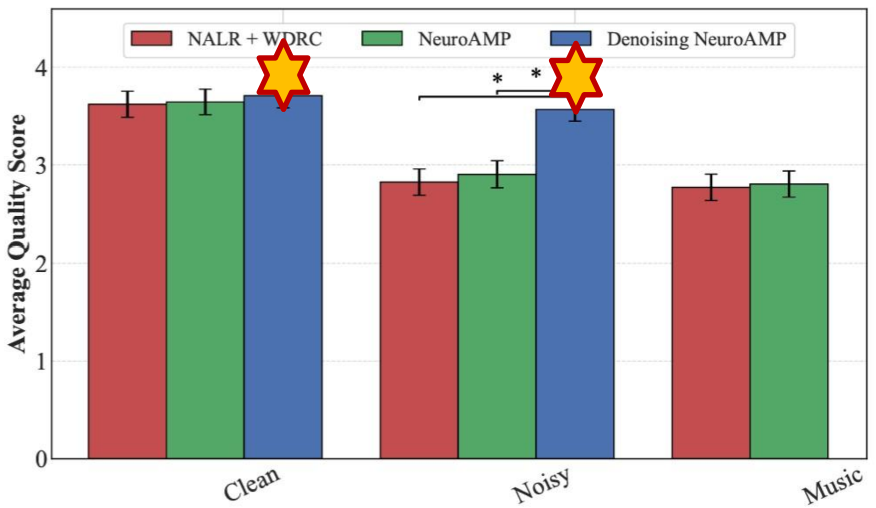


Input Speech (X)

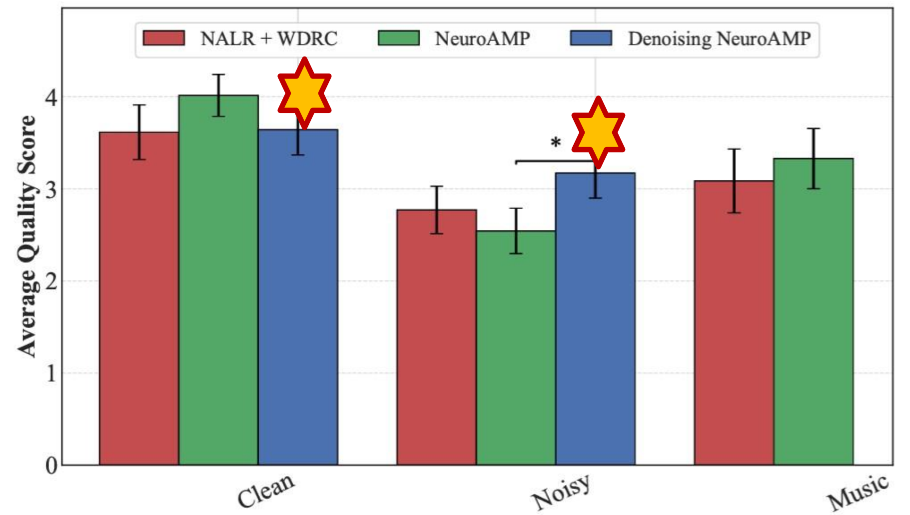
Processed Speech (Ỹ)

Hearing Aids

Simulation Results



Subjective Test Results from HA Users



(1) Simulation Results: The end-to-end HA approach **performs comparably** to conventional HA in **clean** conditions, and **outperforms** conventional HA in **noisy** environments.
(2) Subjective Test Results from HA Users: **Consistent trends** were observed, aligning with the simulation outcomes.

[5] NeuroAMP: A Novel End-to-end General Purpose Deep Neural Amplifier for Personalized Hearing Aids, to appear in IEEE Transactions on Artificial Intelligence.

Related Works

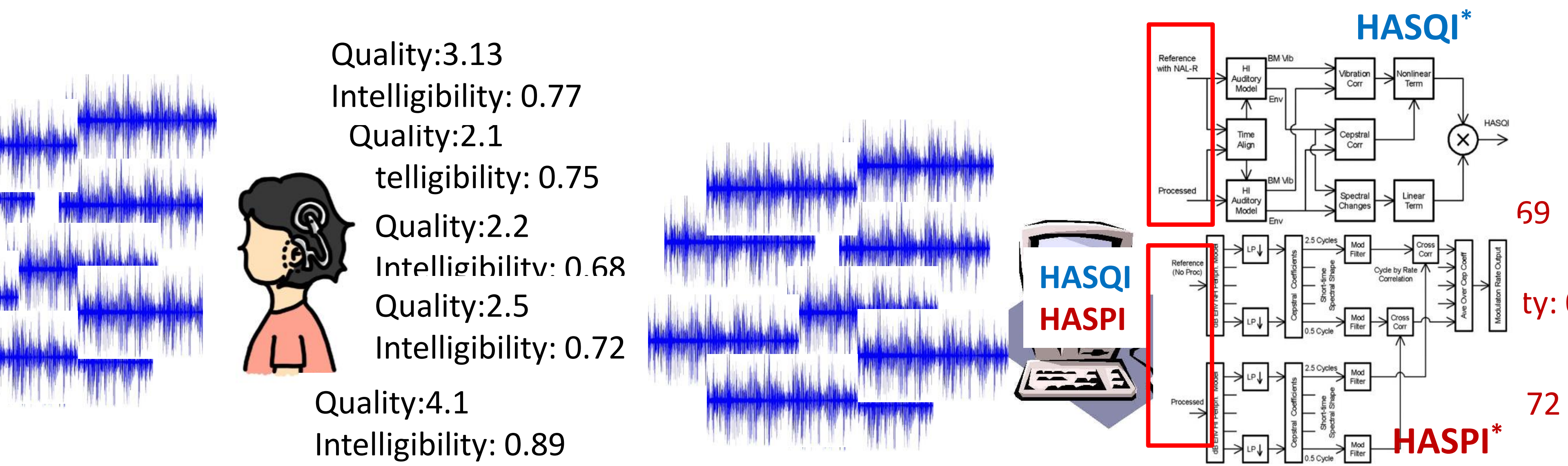
No	Title	System	Method
1	Deep Learning–Based Noise Reduction Approach to Improve Speech Intelligibility for Cochlear Implant Recipients	Cochlear Implant	Speech Enhancement
2	Adversarial Learning For End-to-end Cochlear Speech Denoising Using Lightweight Deep Learning Models	Cochlear Implant	Speech Enhancement
3	Improving the Intelligibility of Speech for Simulated Electric and Acoustic Stimulation Using Fully Convolutional Neural Networks	Cochlear Implant	Speech Enhancement
4	Speech Enhancement Based on Neural Networks Improves Speech Intelligibility in Noise for Cochlear Implant Users	Cochlear Implant	Speech Enhancement
5	Envelope-Based Multichannel Noise Reduction for Cochlear Implant Applications	Cochlear Implant	Speech Enhancement
6	Deep Learning-Based Coding Strategy for Improved Cochlear Implant Speech Perception in Noisy Environments	Cochlear Implant	Coding Strategy
7	A Deep Denoising Sound Coding Strategy for Cochlear Implants	Cochlear Implant	Coding Strategy
8	A Study of Joint Effect on Denoising Techniques and Visual Cues to Improve Speech Intelligibility in Cochlear Implant Simulation	Cochlear Implant	Audio-Visual Integration
9	A Neural-Network Framework for the Design of Individualised Hearing-Loss Compensation	Hearing Aids	DNN (+ Auditory Model)
10	Dynamic Processing Neural Network Architecture for Hearing Loss Compensation	Hearing Aids	DNN (+ Auditory Model)

Related Works

No	Title	System	Method
11	Advances in Intelligent Hearing Aids: Deep Learning Approaches to Selective Noise Cancellation	Hearing Aids	CRN, Transformer
12	Deep Learning Reinvents the Hearing Aid	Hearing Aids	DNN
13	Evaluating Real-World Benefits of Hearing Aids With Deep Neural Network–Based Noise Reduction	Hearing Aids	DNN
14	A U-Net Convolutional Neural Network for Hearing-Aid Speech Enhancement	Hearing Aids	CNN (U-Net)
15	Noise Reduction and Dereverberation for Assisting Hearing-Impaired Listeners Using Recurrent Neural Networks	Hearing Aids	RNN
16	Online Personalization of Compression in Hearing Aids via Maximum Likelihood Inverse Reinforcement Learning	Hearing Aids	Reinforcement Learning (DNN)
17	Speech Denoising and Compensation for Hearing Aids using an FTCRN-based Metric GAN	Hearing Aids	GAN (FTCRN-based)
18	Neural-WDRC: A Deep Learning Wide Dynamic Range Compression Method Combined With Controllable Noise Reduction for Hearing Aids	Hearing Aids	DNN (Two-stage Signal and Noise Separation)
19	Controllable Joint Noise Reduction and Hearing Loss Compensation using a Differentiable Auditory Model	Hearing Aids	DNN (+ Auditory Model)
20	AADNet: An End-to-end Deep Learning Model for Auditory Attention Decoding	Hearing Aids	DNN (+ Auditory Model)

Speech Assessment for Hearing Loss

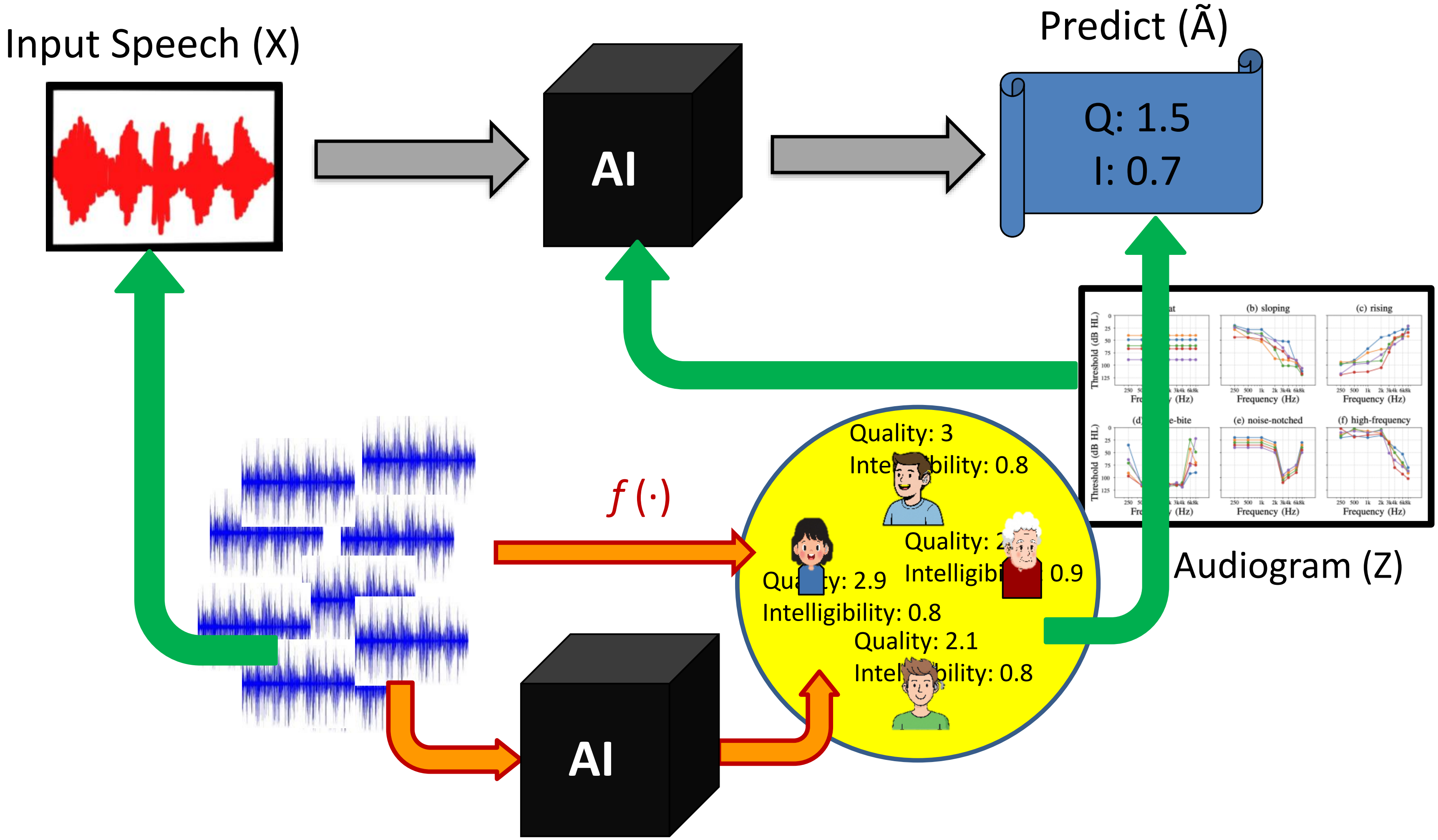
Neural Speech Assessment for Hearing Loss



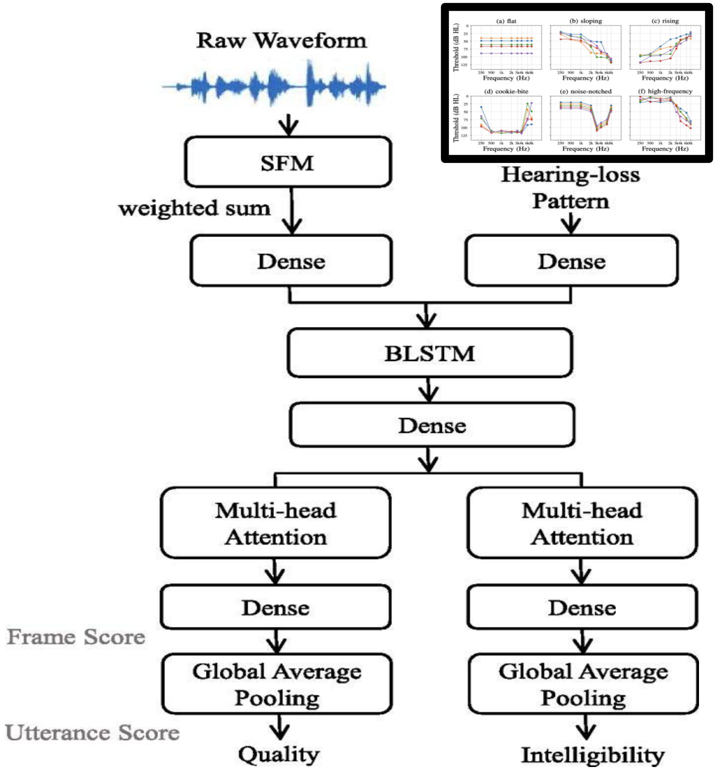
- Subjective functions: **a large scale** of tests is needed to avoid biases.
- Objective function: HASPI and HASQI are popularly used.
- However, both HASQI and HASPI are **intrusive** speech assessment metrics (**clean references** are required.)
- We aim to develop **non-intrusive** speech assessment models.

[6] An Overview of the HASPI and HASQI Metrics for Predicting Speech Intelligibility and Speech Quality for Normal Hearing, Hearing Loss, and Hearing Aids, Hearing Research 2022.

Neural Speech Assessment for Hearing Loss



Neural Speech Assessment for Hearing Loss

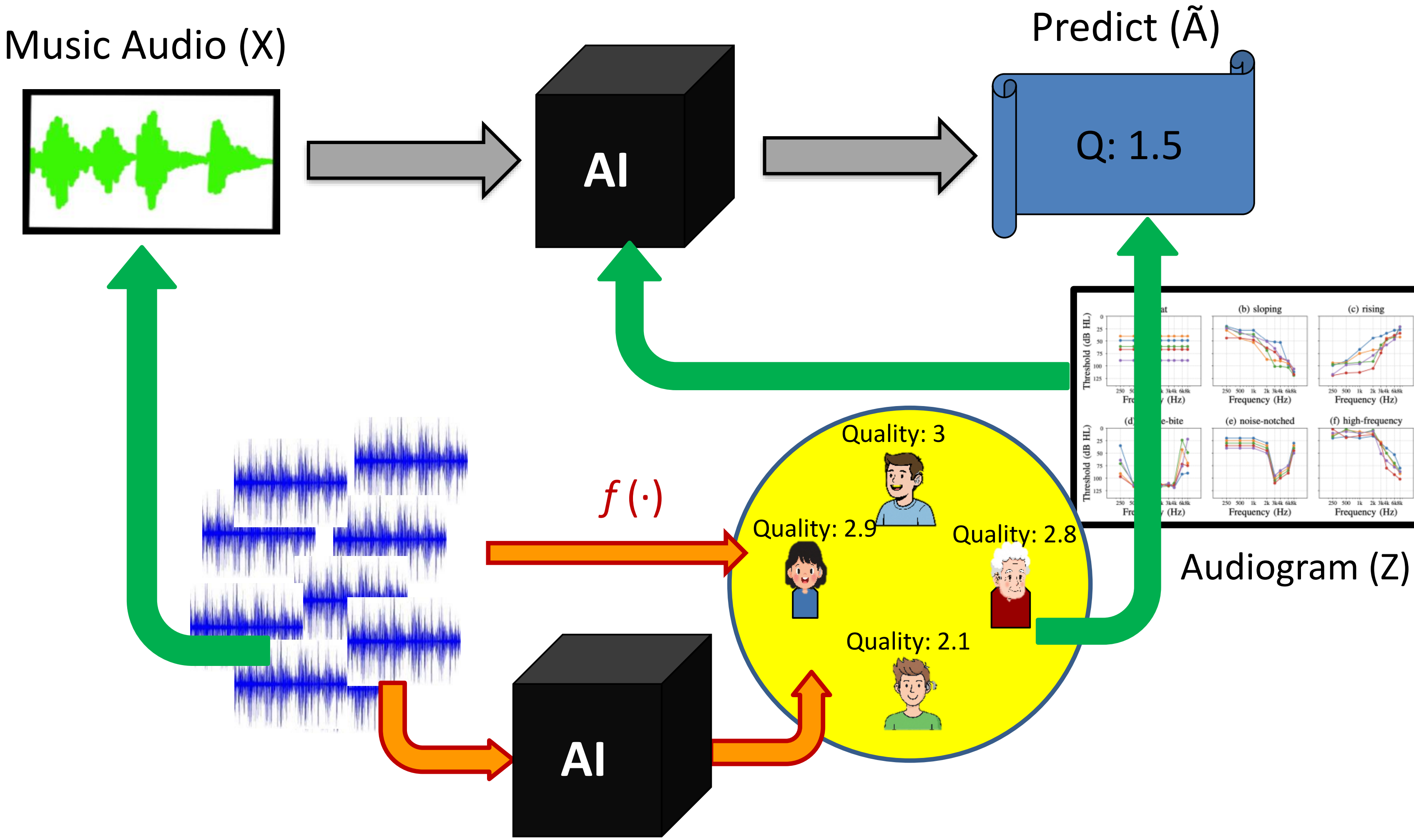


$$L_{Qual} = \frac{1}{B}\sum_{n=1}^{B}\left[\boxed{(\hat{Q}_n - Q_n)^2} + \frac{1}{T_n}\sum_{t=1}^{T_n}\boxed{(\hat{Q}_n - q_{n,t})^2}\right]$$

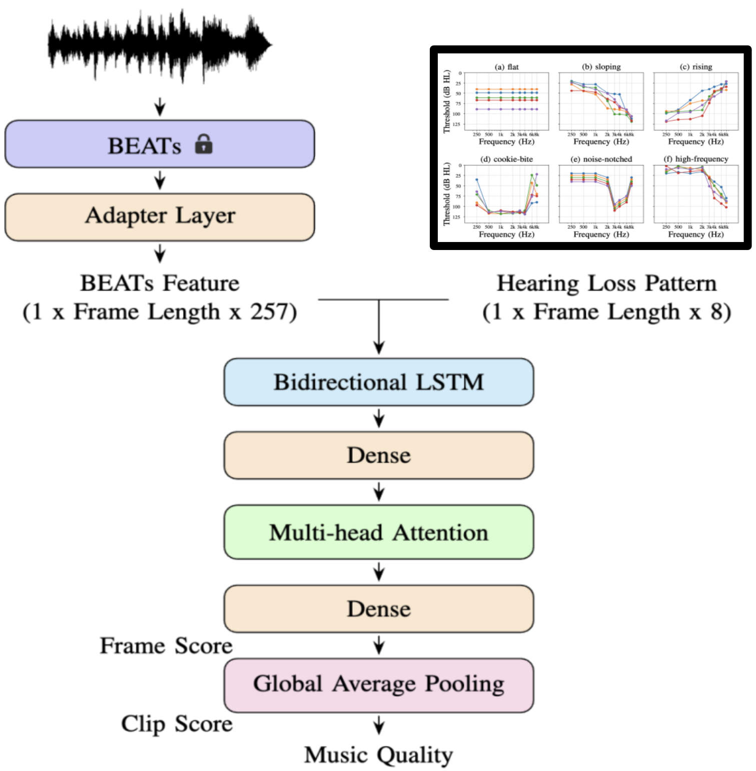
- **Speech Foundation Model Feature Extractor:**
  - Pre-trained Transformer-based models extract **meaningful acoustic representations** from input.
- **Adapter Layer:**
  - A **dense layer** transforms audio features, optimized for assessment.
- **BLSTM:**
  - A **BLSTM** network captures temporal dependencies.
- **Attention Mechanism:**
  - Enhances the model's ability to focus on **important acoustic elements** , improving assessment accuracy.

[7] Multi-objective Non-intrusive Hearing-aid Speech Assessment Model," Journal of the Acoustical Society of America (JASA), volume 156, pages 3574-3587, November 2024.

Neural Music Audio Assessment for Hearing Loss



Neural Music Audio Assessment for Hearing Loss



$$L_{Qual} = \frac{1}{B}\sum_{n=1}^{B}\left[\boxed{(\hat{Q}_n - Q_n)^2} + \frac{1}{T_n}\sum_{t=1}^{T_n}\boxed{(\hat{Q}_n - q_{n,t})^2}\right]$$

- **BEATS Feature Extractor:**
  - Pre-trained Transformer-based models extract **meaningful acoustic representations** from input.
- **Adapter Layer:**
  - A **dense layer** transforms audio features, optimized for assessment.
- **BLSTM:**
  - A **BLSTM** network captures temporal dependencies.
- **Attention Mechanism:**
  - Enhances the model's ability to focus on **important acoustic elements**, improving assessment accuracy.

[8] HAAQI-Net: A Non-intrusive Neural Music Audio Quality Assessment Model for Hearing Aids, IEEE Transactions on Audio, Speech and Language Processing, vol. 33, pp. 1877-1892, 2025.

Related Works

No	Title	System	Method
1	HASA-Net: A Non-Intrusive Hearing-Aid Speech Assessment Network	Hearing Aids	BLSTM
2	MBI-Net: A Non-Intrusive Multi-Branched Speech Intelligibility Prediction Model for Hearing Aids	Hearing Aids	CNN + BLSTM + Attention
3	Non-Intrusive Speech Intelligibility Prediction for Hearing Aids using Whisper and Metadata	Hearing Aids	CNN + BLSTM + Attention (Cross-domain Fea.)
4	Feature Importance across Domains for Improving Non-Intrusive Speech Intelligibility Prediction in Hearing Aids	Hearing Aids	+ BLSTM + Attention (Cross-domain Fea.)
5	Non-Intrusive Speech Intelligibility Prediction Using an Auditory Periphery Model with Hearing Loss	Hearing Aids	2-D CNN (Binaural Data)
6	Speech Intelligibility Prediction Using Binaural Processing for Hearing Loss	Hearing Aids	LSTM+LightGBM
7	Non-Intrusive Binaural Speech Intelligibility Prediction Using MAMBA for Hearing-Impaired Listeners	Hearing Aids	MAMBA
8	A Non-intrusive Speech Quality Evaluation Framework for Hearing Aids Based on Speech Label Assistance and Multi-task Learning Strategy	Hearing Aids	CNN-BiLSTM (+MTL)
9	No Audiogram: Leveraging Existing Scores for Personalized Speech Intelligibility Prediction	Hearing Aids	Support-Based Prediction Module
10	Towards Clinically Feasible Nonintrusive Quality and Intelligibility Indices for Hearing Aids	Hearing Aids	Leveraging ASR+SSL

Dataset and Tools

No	Dataset	System	Task
1	The Clarity Prediction Challenge (1, 2, 3)	Hearing Aids	Speech Assessment
2	The Clarity Enhancement Challenge (1, 2, 3)	Hearing Aids	Speech Enhancement
3	Cadenza Challenge: Music Demixing/Remixing for HAs	Hearing Aids	Music Demixing/Remixing
4	Cadenza Lyric Intelligibility Prediction Challenge	Hearing Aids	Lyric Intelligibility Prediction
5	COG-MHEAR Audio-Visual Speech Enhancement Challenge (1, 2, 3, 4)	Hearing Aids	Audio-Visual Speech Enhancement Challenge

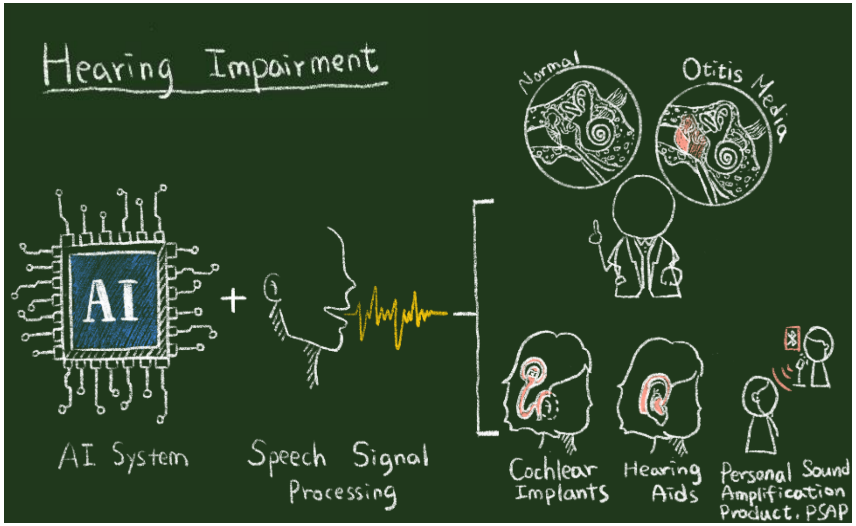
Clarity-2025
One-day workshop
Time: 22nd August, 2025 (9:00 to 15:20)
Location: X Technical University (TU) Delft

AVSEC-4
Half-day afternoon workshop
Time: 16th August, 2025 (12:00 to 17:00)
Location: Technical University (TU) Delft

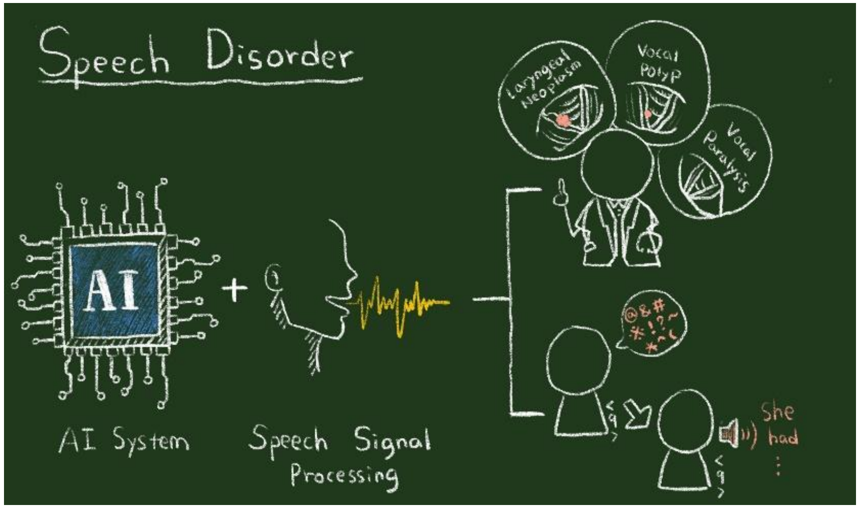
CLIP1
has been accepted as an ICASSP 2026 SP Grand Challenge.

No	Tool	Simulation	Link
1	MSBG (Clarity)	Hearing Aids	https://claritychallenge.org/clarity/clarity.evaluator.msbg.html
2	OpenMHA	Hearing Aids	https://www.openmha.org/
3	Open Speech Platform (OSP)	Hearing Aids	https://github.com/nihospr01/OpenSpeechPlatform
4	Nucleus cochlear implant system	Cochlear Implants	https://github.com/Cochlear-Research/nucleus-toolbox
5	CCi-MOBILE Software Suite	Cochlear Implants	https://crss.utdallas.edu/CILab/CCi-MOBILE.html
6	CISP	Cochlear Implants	https://github.com/Faduma-Ahmed/Cochlear-Implant-signal-processing?utm_source=chatgpt.com

Assistive Oral Communication Technologies



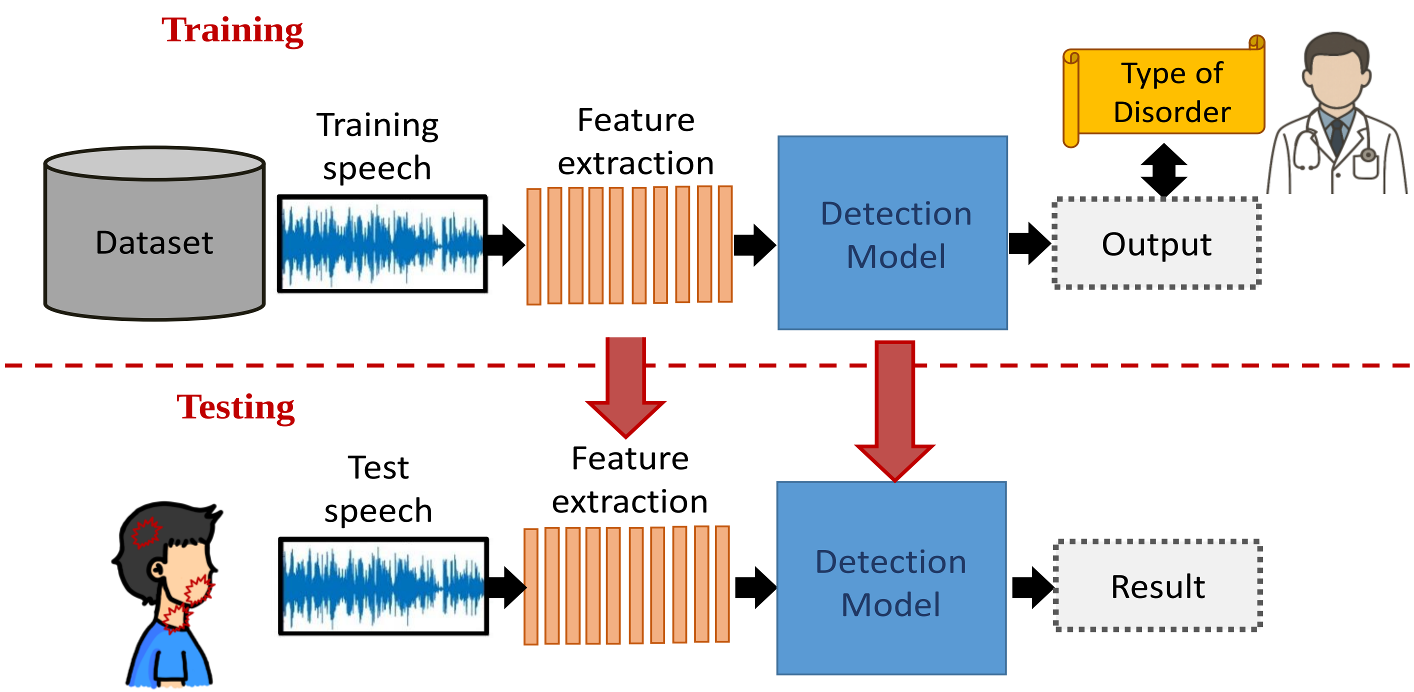
- **Assistive Hearing Technologies**
  - Diagnosis
    - ✓ Automatic otitis media with effusion (OME) detection
  - Speech Enhancement
    - ✓ Speech processing for HA/CI
    - ✓ Multimodal processing for HA/CI
    - ✓ Speech assessment for hearing loss



- **Assistive Speaking Technologies**
  - Diagnosis
    - ✓ Pathological voice detection
    - ✓ Pathological voice assessment
  - Speech Enhancement
    - ✓ Dysarthric speech
    - ✓ After oral surgery
    - ✓ Electrolaryngeal speech

Pathological Voice Detection

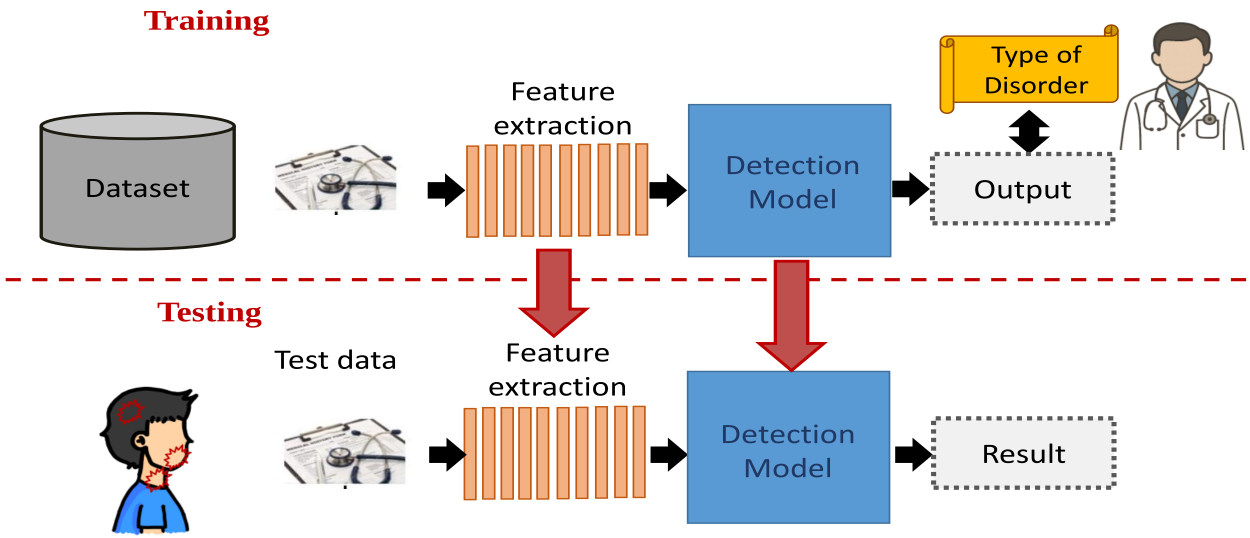
Pathological Voice Detection (Acoustics)



	SVM Accuracy+Std	GMM Accuracy+Std	DNN Accuracy+Std
MFCC	98.28%±2.36%	98.26%±1.80%	99.14%±1.92%
MFCC+Δ	93.04%±2.74%	90.24%±4.18%	94.26%±2.25%
MFCC+Δ+ΔΔ	87.40%±1.92%	90.20±3.83%	90.52%±2.00%

[9] Detection of Pathological Voice using Cepstrum Vectors: A Deep Learning Approach, Journal of Voice, vol. 33 (5), pp. 634-641, 2019.

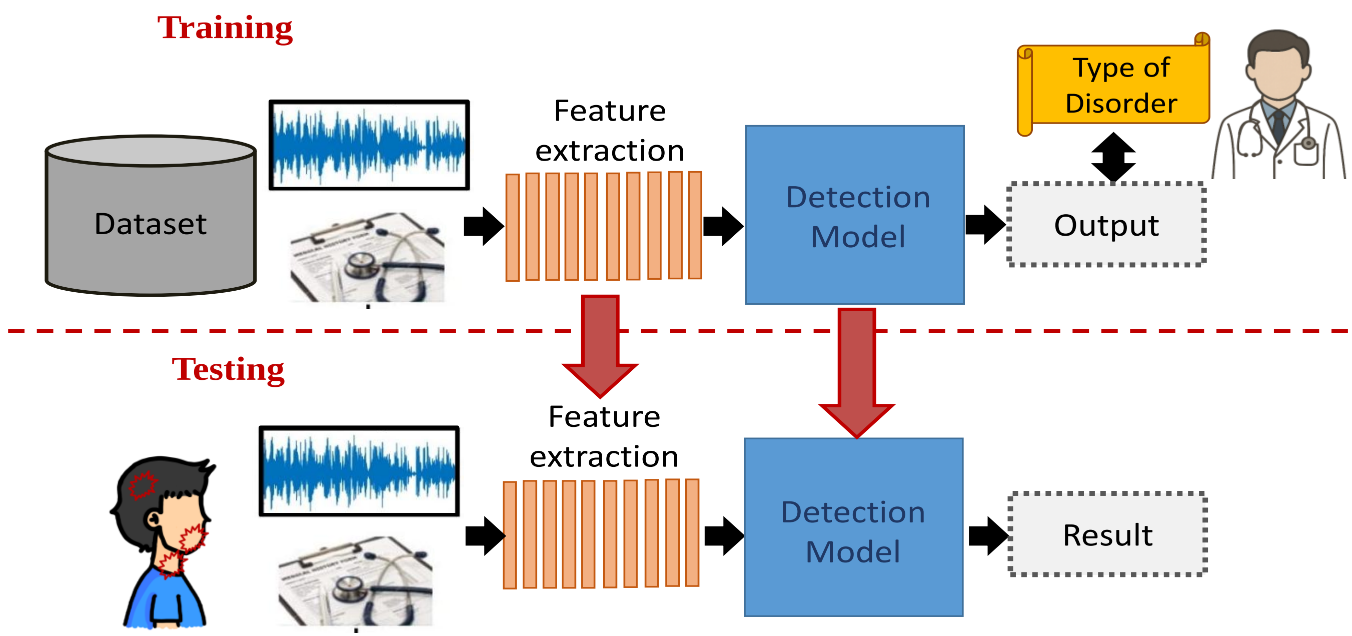
Pathological Voice Detection (Medical Record)



	Neoplasm	Phonotrauma	Palsy	Overall
Demographic				
SVM	73.0%±10.4%	72.8%±2.34%	67.1%±6.99%	71.0%±3.87%
ANN	80.0%±7.07%	71.2%±2.74%	62.6%±4.33%	71.3%±1.85%
Symptomatic				
SVM	60.0%±12.3%	61.4%±6.07%	73.6%±6.20%	65.0%±3.07%
ANN	56.0%±5.48%	68.4%±5.25%	72.3%±5.40%	65.6%±2.29%
Demographic + Symptomatic				
SVM	82.0%±7.58%	79.5%±1.45%	83.9%±7.90%	81.8%±4.87%
ANN	83.0%±7.58%	79.4%±1.83%	86.5%±2.70%	83.0%±1.58%

[10] Demographic and Symptomatic Features of Voice Disorders and Their Potential Application in Classification using Machine Learning Algorithms, Folia Phon. et Logo. 70, 2018.

Pathological Voice Detection (Multimodal)

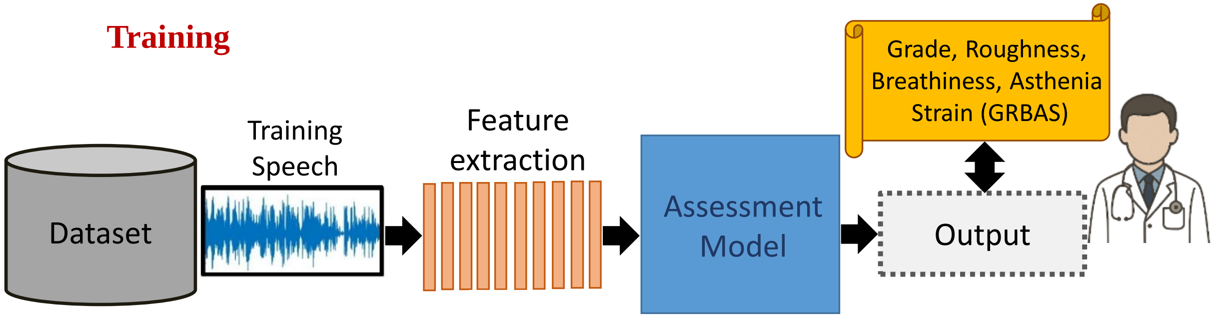


	Neoplasm	Phonotrauma	Vocal Palsy	Accuracy	UAR
	Sensitivity (Recall)				
Acoustic Signals	63.00%±17.89%	95.36%±4.39%	34.40%±20.12%	76.94%±6.71%	64.25%±11.04%
Medical Record	59.00%±11.40%	91.54%±3.67%	70.40%±2.19%	81.56%±1.25%	73.65%±3.49%
Multimodal	79.00%±14.75%	95.36±3.03%	70.40%±10.43%	87.26%±2.23%	81.59%±5.94%

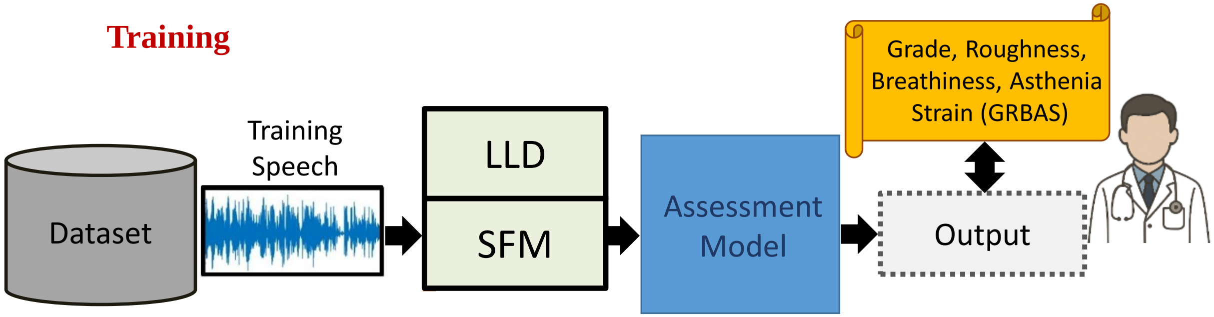
[11] Combining Acoustic Signals and Medical Records to Improve Pathological Voice Classification, APSIPA Transactions on Signal and Information Processing, vol. 8, 2019.

Pathological Voice Assessment

Pathological Voice Assessment

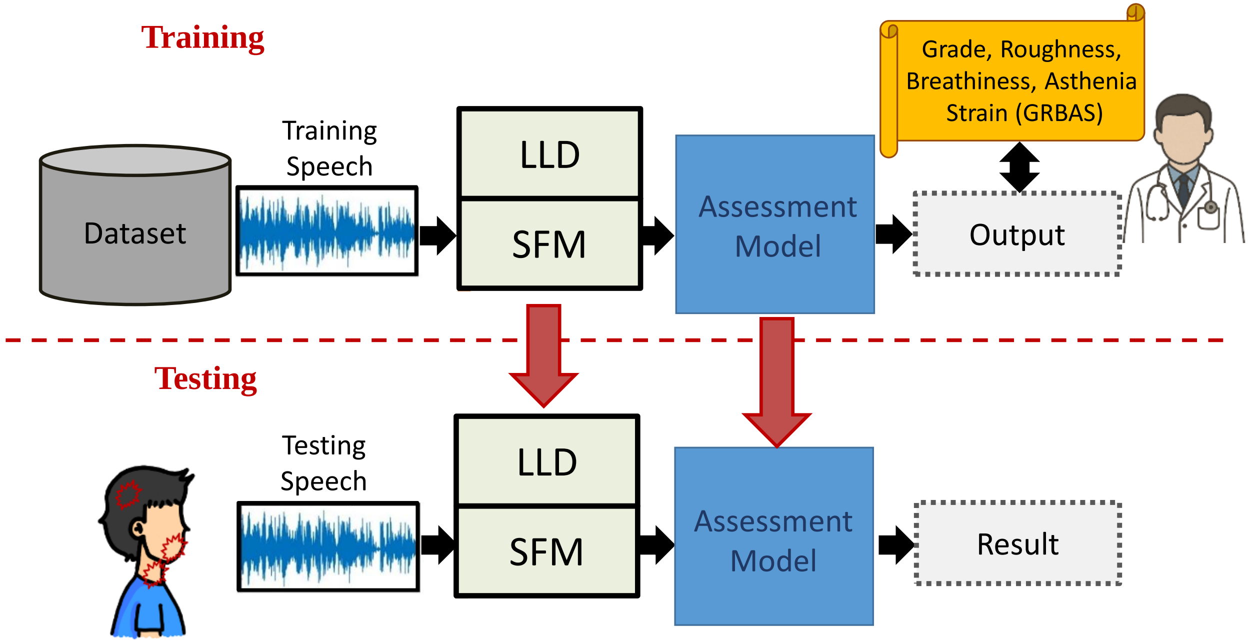


Pathological Voice Assessment



[12] Towards Robust Assessment of Pathological Voices via Combined Low-Level Descriptors and Foundation Model Representations. arXiv preprint arXiv:2505.21356, 2025.

Pathological Voice Assessment



Features	PVQD-A		PVQD-S	
	RMSE	PCC	RMSE	PCC
LLD	14.76	0.642	-	-
SFM (WavLM)	9.891	0.865	9.209	0.870
SFM+LLD	8.594	0.877	8.720	0.883

[12] Towards Robust Assessment of Pathological Voices via Combined Low-Level Descriptors and Foundation Model Representations. arXiv preprint arXiv:2505.21356, 2025.

Related Works

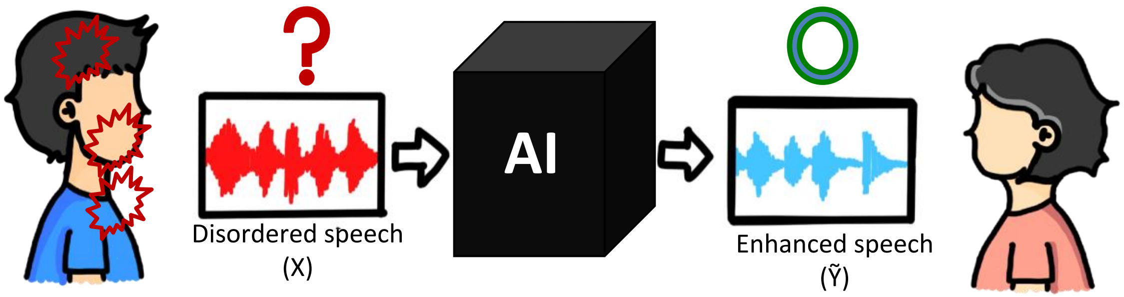
No	Title	System	Method
1	Continuous Speech for Improved Learning Pathological Voice Disorders	Pathological Voice	DNN
2	A Voice Disease Detection Method Based on MFCCs and Shallow CNN	Pathological Voice	CNN
3	MVP: Multi-source Voice Pathology Detection	Pathological Voice	Transformer (+Pretrained Model)
4	Voice Pathology Detection and Classification using Online Sequential Extreme Learning Machine (OSELM)	Pathological Voice	Extreme Learning Machine
5	Voice Pathology Detection using Convolutional Neural Networks with Electroglottographic (EGG) and Speech Signals	Pathological Voice	CNN
6	Voice Disorder Detection using LSTM Model	Pathological Voice	LSTM
7	Pathological Voice Classification using Mel-Cepstrum Vectors and SVM	Pathological Voice	SVM
8	Multi-Dimensional Features Extraction for Voice Pathology Recognition using ResNet	Pathological Voice	ResNet
9	Learning Strategies for Voice Disorder Detection	Pathological Voice	Machine Learning
10	Evaluation of Phone Posterior Probabilities for Pathology Detection in Speech Data using Deep Learning Models	Pathological Voice	CNN (+Pretrained Model)

Related Works

No	Title	System	Method
1	AROA based Pre-Trained Model of Convolutional Neural Network for Voice Pathology Detection and Classification	Pathological Voice	CNN (+AROA pretraining)
2	Voice Pathology Detection and Classification Using Convolutional Neural Network Model	Pathological Voice	CNN
3	Attentive-based Multi-Level Feature Fusion for Voice Disorder Diagnosis	Parkinson Voice	ECAPA-TDNN (+ Pretrained Models)
4	Deep Neural Networks for Voice Quality Assessment based on the GRBAS Scale	Pathological Voice	DNN
5	Emulating Perceptual Evaluation of Voice Using Scattering Transform Based Features	Pathological Voice	Scattering Transform
6	Lightly Weighted Automatic Audio Parameter Extraction for the Quality Assessment of Consensus Auditory-perceptual Evaluation of Voice	Pathological Voice	Traditional Machine Learning (+LLD)
7	Bilingual Dual-Head Deep Model for Parkinson's Disease Detection from Speech	Parkinson's Disease	SSL + Wavelet
8	Clinical Examination of Voice	Pathological Voice	Protocol
9	French-language Adaptation of the Consensus Auditory-Perceptual Evaluation of Voice (CAPEV)	Pathological Voice	Protocol
10	Crowdsourced Perceptual Ratings of Voice Quality in People With Parkinson's Disease Before and After Intensive Voice and Articulation Therapies: Secondary Outcome of A Randomized Controlled Trial	Parkinson Voice	Protocol

Speech Enhancement for Speaking Disorders

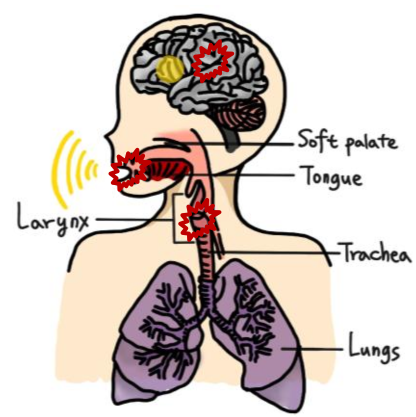
Speech Enhancement for Speaking Disorder



- Solutions

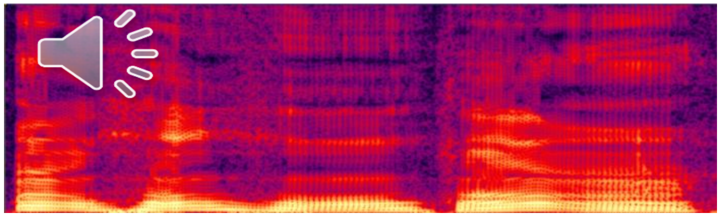
	Voice conversion/speech enhancement	Automatic speech recognition (ASR) + text-to-speech (TTS)
Pros	Low computation cost (edge device)	Leveraging advanced systems
Cons	Relatively lower performance	High computation cost

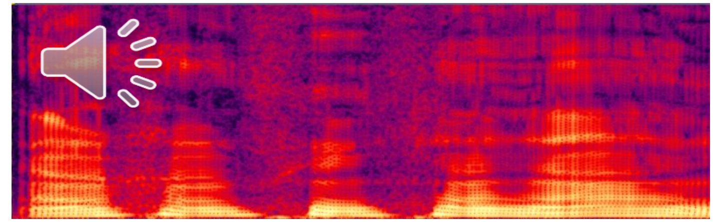
- **Oral Surgical** speech enhancement
- **Dysarthric** speech enhancement
- **Electrolaryngeal** speech enhancement



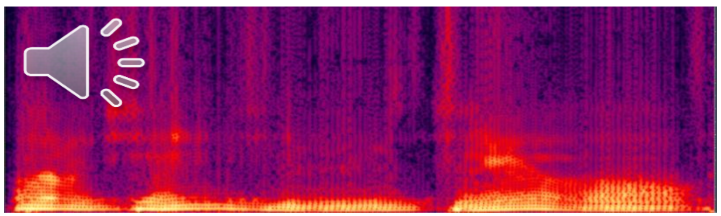
Oral Surgical Speech Enhancement

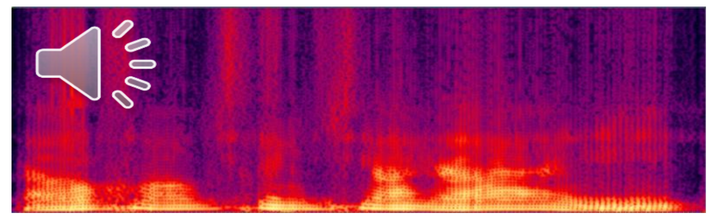
Original:





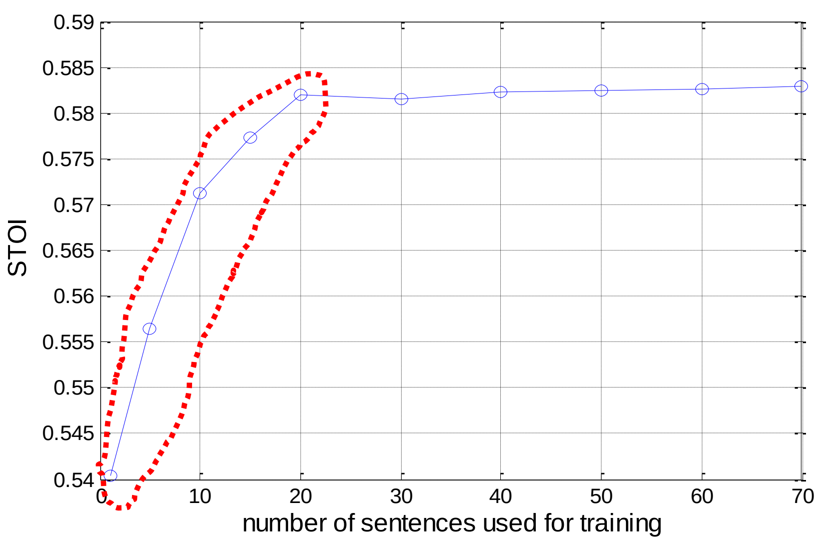
After Conversion:





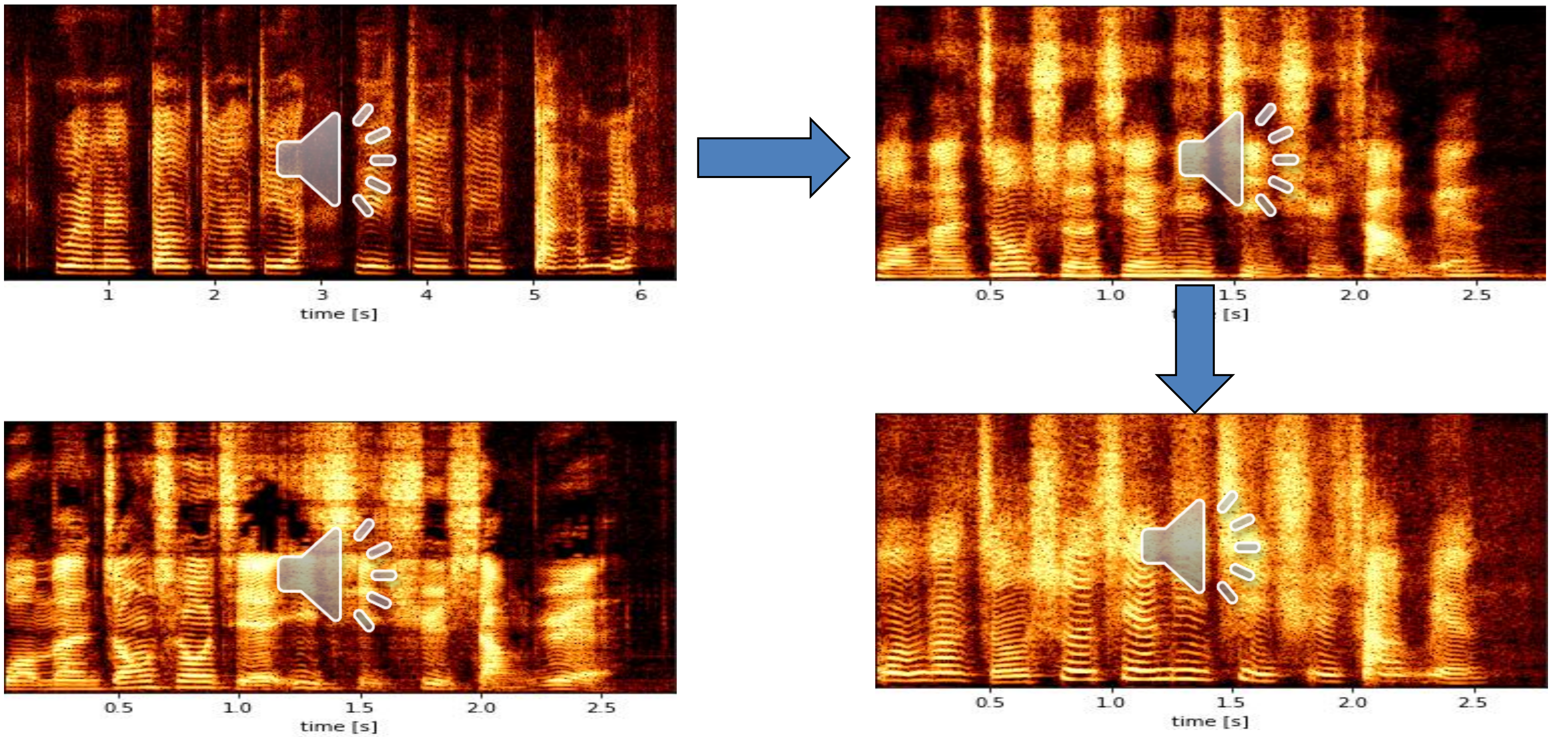
衛生紙給我
wei4 sheng1 zhi3 gei3 wo3

遙控器在哪裡
yao2 kong4 qi4 zai4 na3 li3



[13] Joint Dictionary Learning-Based Non-Negative Matrix Factorization for Voice Conversion to Improve Speech Intelligibility After Oral Surgery, IEEE Transactions on Biomedical Engineering vol. 64 (11), pp. 2584-2594, 2016

Dysarthric Speech Enhancement



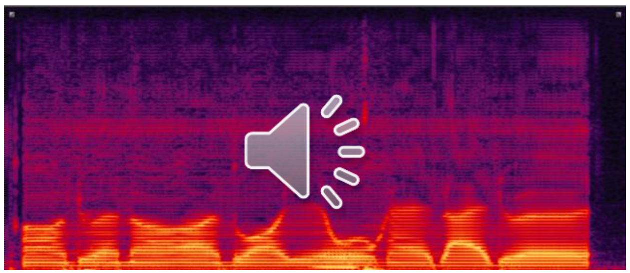
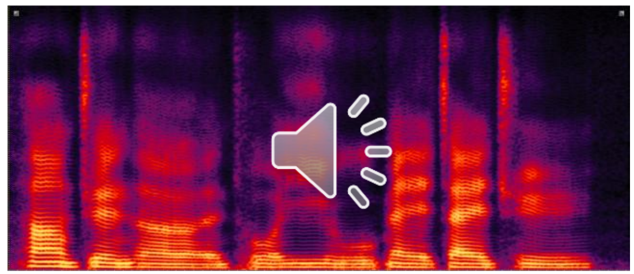
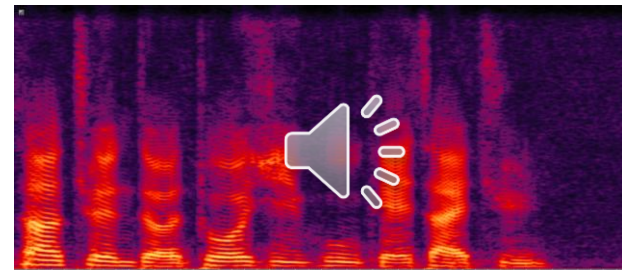
我們中秋節一起去賞月
(wo3 men5 zhong1 qiou1 jie2 i4 qi3 qy4 shang3 yue4)

[14] A Preliminary Study of a Two-Stage Paradigm for Preserving Speaker Identity in Dysarthric Voice Conversion," Interspeech 2021.

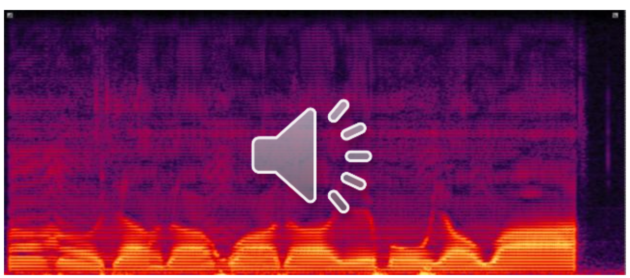
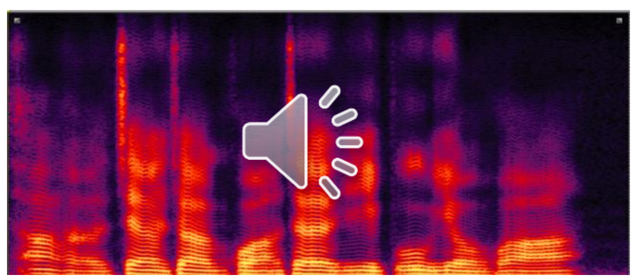
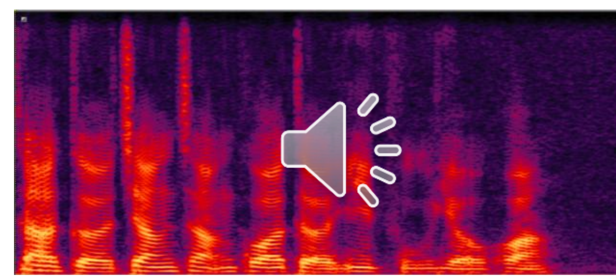
Electrolaryngeal Speech Enhancement

Original | MT-CLDNN | Seq2seq ELVC

Sample 1: 他捐了很多衣物給災區 (ta1 juan1 le5 hen3 duo1 i1 wu4 gei3 zai1 qv1)

Sample 2: 那個牆上掛著一幅油畫 (na4 ge5 qiang2 shang4 gua4 zhe5 i1 fu2 you2 hua4)

[15] Mandarin Electrolaryngeal Speech Voice Conversion with Sequence-to-sequence Modeling, in Proc. ASRU 2021.

Related Works

No	Title	Disorder Type	Method
1	Pretraining and Adaptation Techniques for Electrolaryngeal Speech Recognition	Electrolaryngeal Speech	Adapting ASR
2	Robust Sequence-to-sequence Voice Conversion for Electrolaryngeal Speech Enhancement in Noisy and Reverberant Conditions	Electrolaryngeal Speech	Robust Voice Conversion
3	Resolving Domain Mismatches in Electrolaryngeal Speech Enhancement With Linguistic Intermediates	Electrolaryngeal Speech	Speech Enhancement (+Linguist)
4	An Investigation of Fundamental Frequency Pattern Prediction for Japanese Electrolaryngeal Speech Enhancement Based on Frame-Wise Phoneme Representations	Electrolaryngeal Speech	Speech Enhancement (+Phonetic)
5	Exemplar-Based Methods for Mandarin Electrolaryngeal Speech Voice Conversion	Electrolaryngeal Speech	Voice Conversion (LLE)
6	Mandarin Electrolaryngeal Speech Voice Conversion with Speech Encoder Loss Learning and Seq2seq Modeling	Electrolaryngeal Speech	Voice Conversion (S2S)
7	Electrolaryngeal Speech Intelligibility Enhancement Through Robust Linguistic Encoders	Electrolaryngeal Speech	Voice Conversion (+Linguist)
8	Two-Stage Training Method for Japanese Electrolaryngeal Speech Enhancement Based on Sequence-to-sequence Voice Conversion	Electrolaryngeal Speech	Speech Enhancement (S2S)
9	An Approach for Speech Enhancement with Dysarthric Speech Recognition using Optimization based Machine Learning Frameworks	Electrolaryngeal Speech	Speech Enhancement (+ASR)
10	Collaborative AI Dysarthric Speech Recognition System With Data Augmentation Using Generative Adversarial Neural Network	Electrolaryngeal Speech	ASR (+Data Augmentation)

Related Works

No	Title	Disorder Type	Method
11	Audio-Visual Mandarin Electrolaryngeal Speech Voice Conversion	Electrolaryngeal Speech	Voice Conversion (Audio+Visual)
12	Enhancing Dysarthric Voice Conversion with Fuzzy Expectation Maximization in Diffusion Models for Phoneme Prediction	Dysarthric Speech	Voice Conversion (+Phonetic)
13	Enhancing Dysarthric Speech Recognition Through SepFormer and Hierarchical Attention Network Models with Multistage Transfer Learning	Dysarthric Speech	Speech Enhancement +ASR
14	Dysarthric Speech Transformer: A Sequence-to-sequence Dysarthric Speech Recognition System	Dysarthric Speech	ASR (S2S)
15	Exploring Generative Error Correction for Dysarthric Speech Recognition	Dysarthric Speech	ASR (+Error Correction)
16	On Using the UA-Speech and Torgo Databases to Validate Automatic Dysarthric Speech Classification Approaches	Dysarthric Speech	ASR
17	Speech Vision: An End-to-end Deep Learning-Based Dysarthric Automatic Speech Recognition System	Dysarthric Speech	E2E ASR
18	Improving Speech Enhancement Performance by Leveraging Contextual Broad Phonetic Class Information	Dysarthric Speech	Speech Enhancement (+Phonetic)
19	Voice Conversion for Stuttered Speech, Instruments, Unseen Languages and Textually Described Voices	Shuttered Speech	Voice Conversion (kNN)
20	Parrotron: An End-to-End Speech-to-Speech Conversion Model and its Applications to Hearing-Impaired Speech and Speech Separation	Hearing-Impaired Speech	Voice Conversion

Summary and Future Directions

Summary

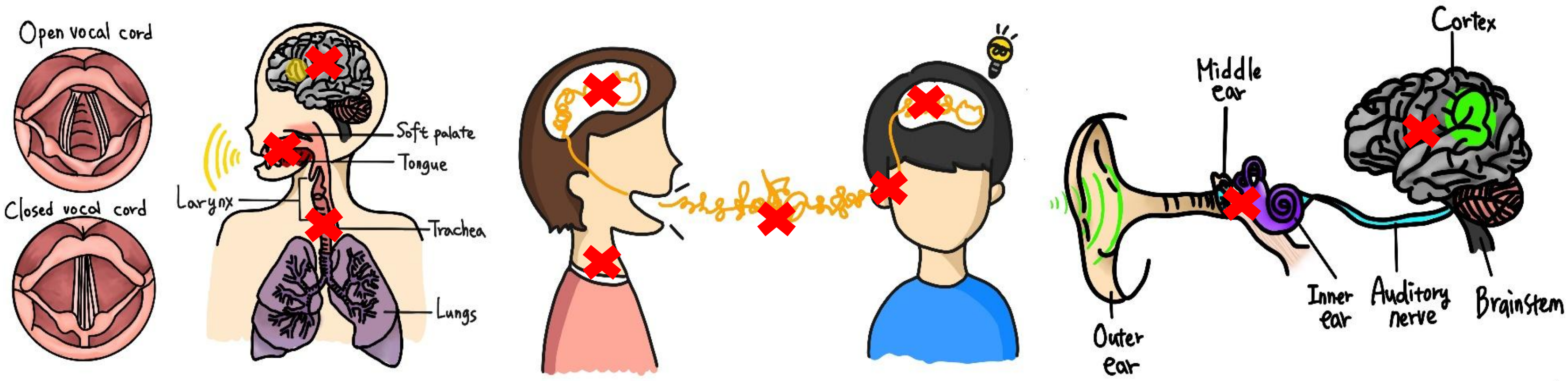
Observations

- Speech plays a fundamental role in **human-to-human** communication.
- Although AI has achieved **remarkable progress** in fields such as multimedia processing and medicine, assistive hearing and speaking technologies are still largely limited by **traditional approaches**.
- Advancing assistive oral communication technologies has the potential to directly **improve the well-being** of individuals with communication challenges.

Research Topics

- Leveraging advanced **multimodal pretrained** models.
- Harness multimodal large language models (**LLMs**).
- Design more effective and adaptive **agentic AI systems.**
- Explore brain responses as **unbiased biomarkers**.

Future Outlook: Toward Barrier Free Oral Communication



Issues to address

- **Dynamic** speech enhancement systems for real-world conditions.
- **Multimodal** assistive speaking and hearing technologies (smart glasses).
- Encoding and decoding speech signals into **brain responses.**
- Leveraging **brain signals** to enhance speech comprehension.

References

[1] Detection of Otitis Media With Effusion Using In-Ear Microphones and Machine Learning, IEEE Sensors Journal, vol. 23, no. 22, pp. 28411-28420, 15 Nov.15, 2023.
[2] A Deep Denoising Autoencoder Approach to Improving the Intelligibility of Vocoded Speech in Cochlear Implant Simulation, IEEE Transactions on Biomedical Engineering, vol. 64 (7), pp. 1568-1578, July 2017.
[3] ElectrodeNet–A Deep-learning-based Sound Coding Strategy for Cochlear Implants,“ IEEE Transactions on Cognitive and Developmental Systems, vol. 16 (1), pp. 346-357, February 2024.
[4] End-to-End Audio-Visual Learning for Cochlear Implant Sound Coding in Noisy Environments, under review.
[5] NeuroAMP: A Novel End-to-end General Purpose Deep Neural Amplifier for Personalized Hearing Aids, to appear in IEEE Transactions on Artificial Intelligence.
[6] An Overview of the HASPI and HASQI Metrics for Predicting Speech Intelligibility and Speech Quality for Normal Hearing, Hearing Loss, and Hearing Aids, Hearing Research 2022.
[7] Multi-objective Non-intrusive Hearing-aid Speech Assessment Model," Journal of the Acoustical Society of America (JASA), volume 156, pages 3574-3587, November 2024.
[8] HAAQI-Net: A Non-intrusive Neural Music Audio Quality Assessment Model for Hearing Aids, IEEE Transactions on Audio, Speech and Language Processing, vol. 33, pp. 1877-1892, 2025.
[9] Detection of Pathological Voice using Cepstrum Vectors: A Deep Learning Approach, Journal of Voice, vol. 33 (5), pp. 634-641, 2019.
[10] Demographic and Symptomatic Features of Voice Disorders and Their Potential Application in Classification using Machine Learning Algorithms, Folia Phon. et Logo. 70, 2018.
[11] Combining Acoustic Signals and Medical Records to Improve Pathological Voice Classification, APSIPA Transactions on Signal and Information Processing, vol. 8, 2019.
[12] Towards Robust Assessment of Pathological Voices via Combined Low-Level Descriptors and Foundation Model Representations. arXiv preprint arXiv:2505.21356, 2025.
[13] Joint Dictionary Learning-Based Non-Negative Matrix Factorization for Voice Conversion to Improve Speech Intelligibility After Oral Surgery, IEEE Transactions on Biomedical Engineering vol. 64 (11), pp. 2584-2594, 2016
[14] A Preliminary Study of a Two-Stage Paradigm for Preserving Speaker Identity in Dysarthric Voice Conversion," Interspeech 2021.
[15] Mandarin Electrolaryngeal Speech Voice Conversion with Sequence-to-sequence Modeling, in Proc. ASRU 2021.

Thank You Very Much for Your Attention