The 2nd Clarity Workshop on Machine Learning Challenges for Hearing Aids
June 29, 2022

Clarity

OBISHI: Objective Binaural Intelligibility Score for the Hearing Impaired

Candy Olivia Mawalim, Benita Angela Titalim, Masashi Unoki, Shogo Okada

Japan Advanced Institute of Science and Technology

JAIST
JAPAN
ADVANCED INSTITUTE OF
SCIENCE AND TECHNOLOGY
1990

Outline

1. Introduction
2. Speech Intelligibility Model
3. Proposed Method
4. Evaluation
5. Conclusion

1. Introduction

Research Background

Living environment

People with hearing impairment

Hearing aids

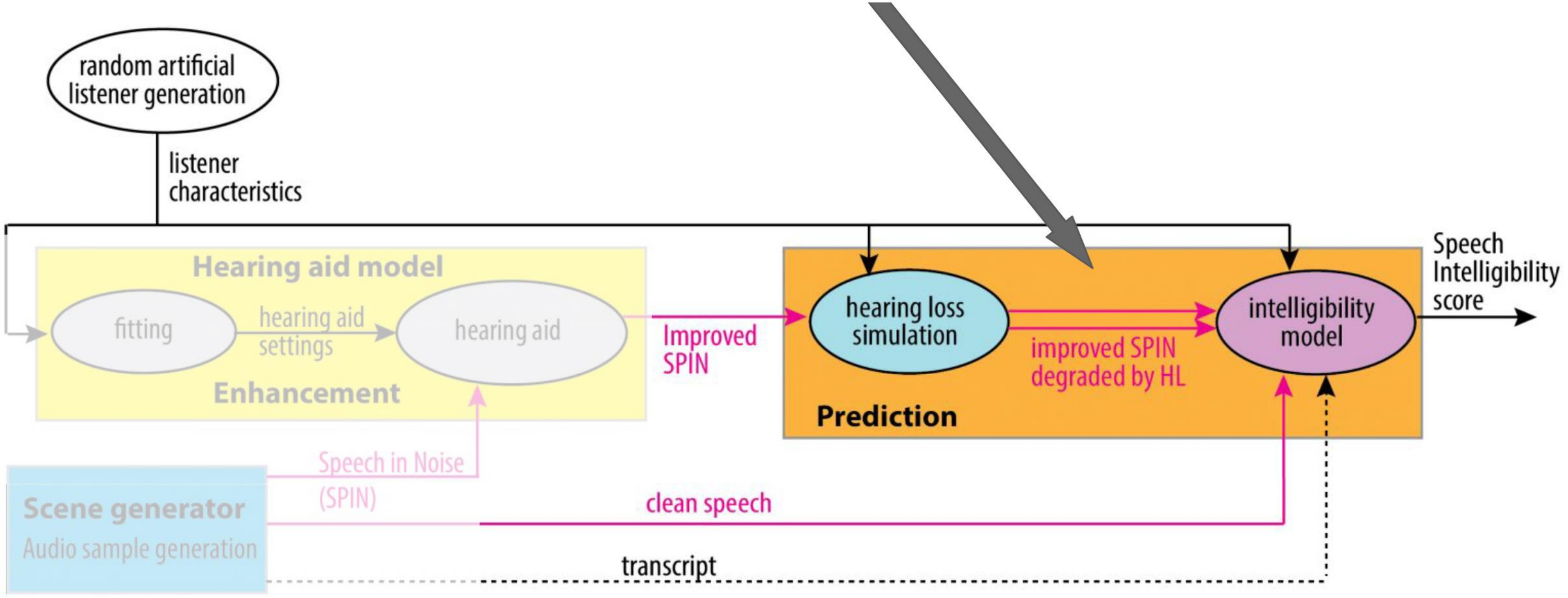
Speech Intelligibility Metrics for Hearing Aids

- Speech intelligibility metrics are essential for hearing aids development.
- Speech intelligibility (SI) often refers to how accurately speech is understood or to the percentage of the number of words the listener correctly identifies.

$$SI = \frac{\#words\ that\ are\ correctly\ identified}{\#words\ in\ sentence}$$

The First Clarity Prediction Challenge (CPC1)

- Main task of the CPC1 is to predict the speech intelligibility of HI listeners when they perceive noisy speech processed by a hearing aid system.



https://claritychallenge.github.io/clarity_CPC1_doc/docs/intro

2. Speech Intelligibility Model

Hearing Aid Speech Prediction Index (HASPI)

- The HASPI by Kates and Arehart is often considered in developing hearing aids as an objective speech intelligibility index.
- The HASPI model includes a comparison of the temporal amplitude envelope (TAE) and temporal fine structure (TFS) that improved the prediction accuracy in both normal hearing (NH) and hearing-impaired (HI) processing.
- Unfortunately, the HASPI model has several drawbacks:
  i. Evaluations are limited to the conditions provided in the training data
  ii. Handle monaural listening
  iii. Only consider the audiogram for the listener's hearing characteristics
  iv. Invalid for tonal language

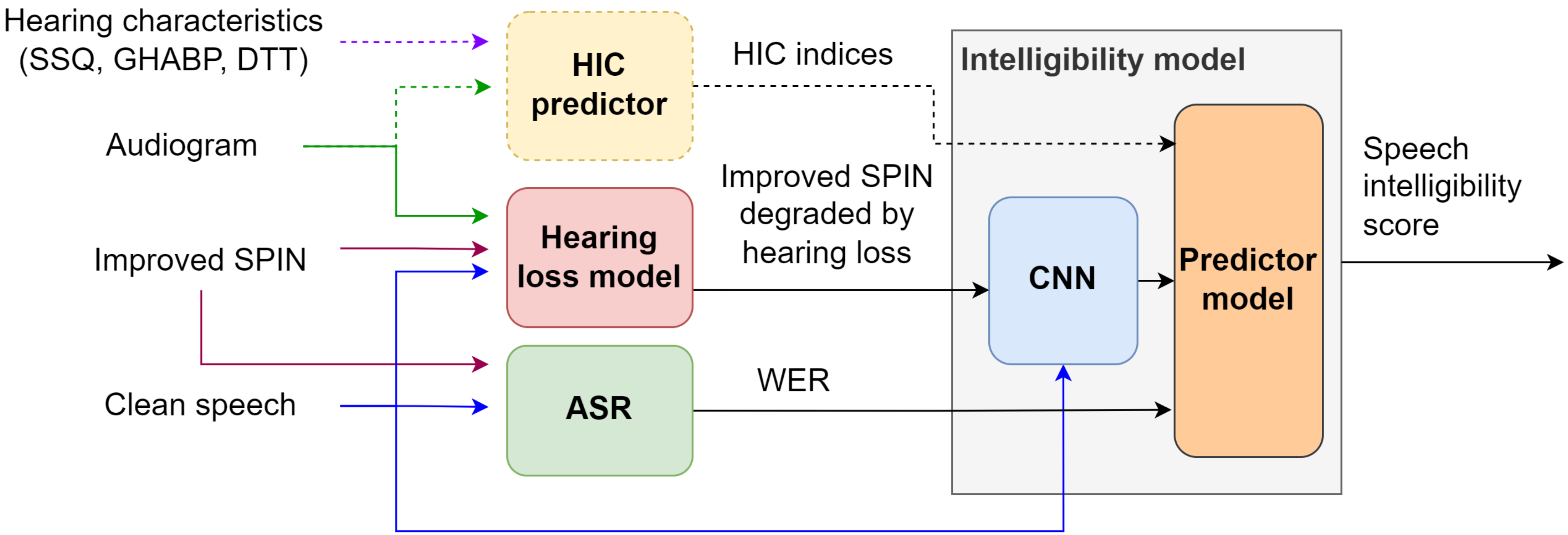
2. Speech Intelligibility Model

MBSTOI (Baseline)

- MBSTOI: Modified Binaural Short-Time Objective Intelligibility
- This model was developed based on the STOI metric and is an extended model of discrete binaural STOI (DBSTOI).
- The MBSTOI generates more accurate predictions than the DBSTOI because it overcomes the tendency of overestimation when the interferers are spatially distributed.
- The MBSTOI model utilized the Cambridge hearing loss model to approximate the HI auditory thresholds by adding internal noise and by attenuating the signals. Thus, the baseline model is sensitive to the level of the processed signal.

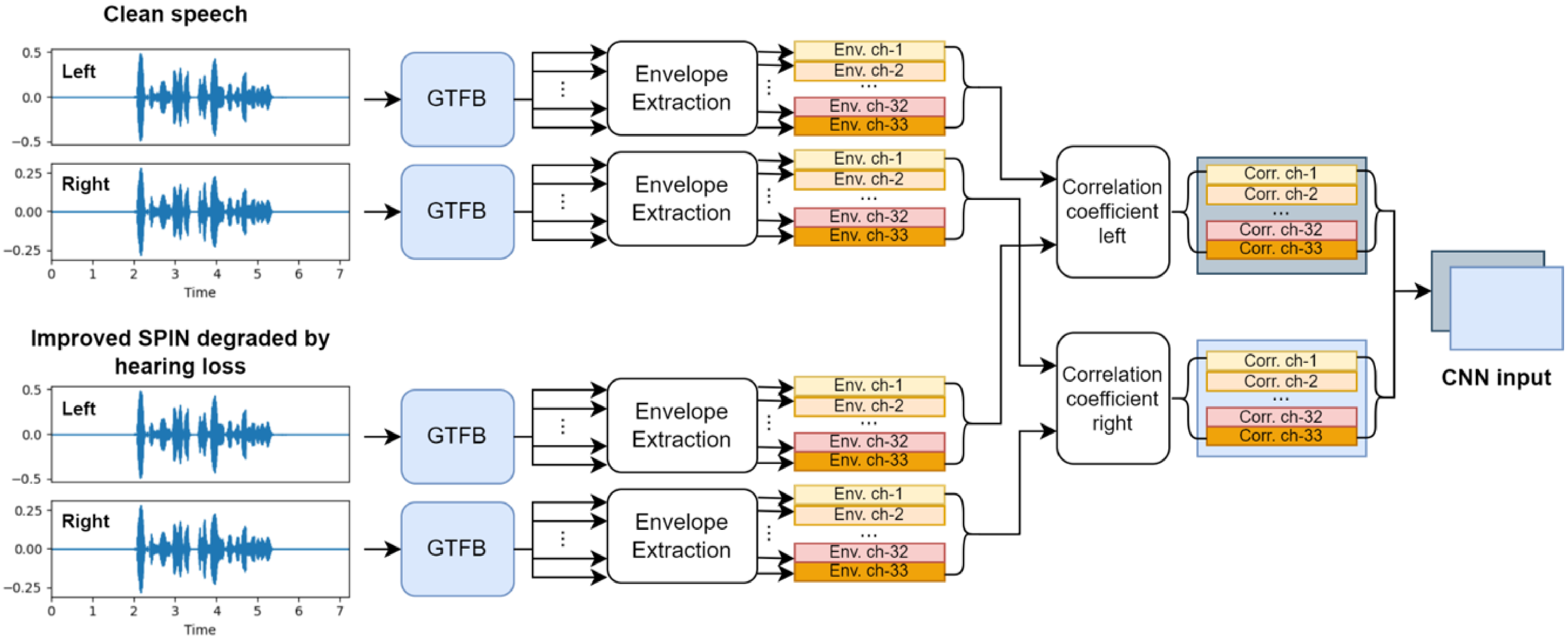
3. Proposed Method

Block Diagram of OBISHI

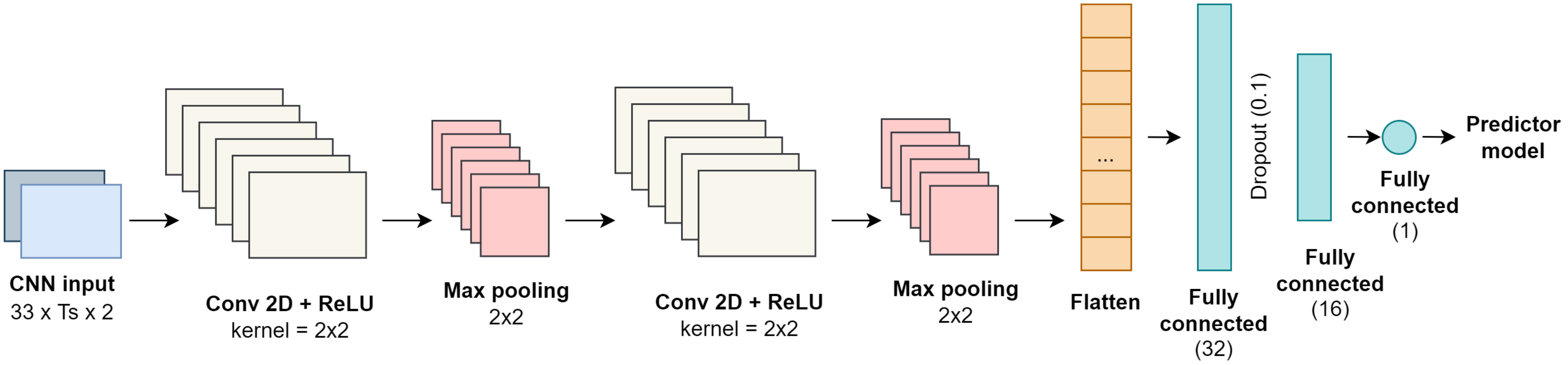


3. Proposed Method

CNN Input Generation



3. Proposed Method
CNN Architecture



4. Evaluation

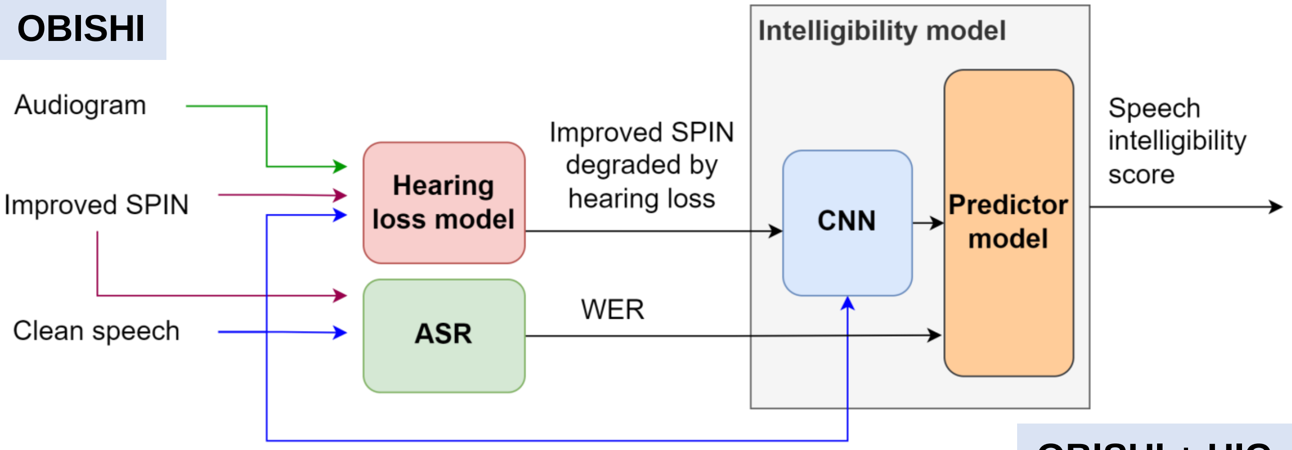
Dataset and Metrics

- **The CPC1 dataset:** *https://claritychallenge.github.io/clarity_CPC1_doc/docs/cpc1_data*
  - It contains wav files: generated scenes, interferers, original target speech spoken by British English speakers, and improved SPINs (the output of SPINs after passing through hearing aid (HA) processors).
  - The metadata provides information related to the scenes, listeners, and transcripts.
  - It has 6 speakers, 10 HA processors, and 27 HI listeners.
  - HI Listeners characteristics: audiogram, SSQ, GHABP, and DTT
  - Two tracks: close-set, open-set (with unseen listener(s) and an unseen HA system)
- **Evaluation Metrics:** Pearson correlations (ρ), root mean square error (RMSE), F1 score (F1), and area under the curve (AUC)

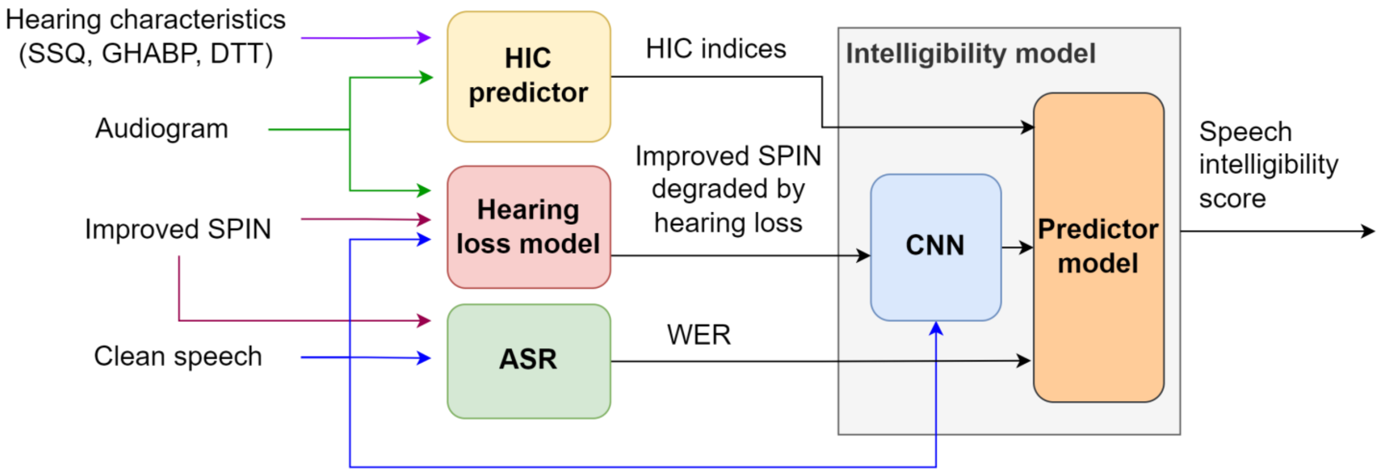
4. Evaluation

Proposed Method

OBISHI



OBISHI + HIC



Results (Development Phase)

- **Close-set track**

  Training/development (train): 4,863 (90% training, 10% development)

- **Open-set track**

  Training/development (train-indep): 3,580
  (unseen 1 listener and 1 system: 647 for development; the rest 2,933 for training)

Method	Track 1 (close-set)				Track 2 (open-set)			
	ρ	RMSE	F1 (%)	AUC (%)	ρ	RMSE	F1 (%)	AUC (%)
Baseline	0.63	33.65 ± 1.42	81.01	76.11	0.48	33.77 ± 0.92	84.57	67.18
HASPI (left)	0.67	36.07 ± 1.34	73.13	71.91	0.43	43.58 ± 1.02	52.91	58.15
HASPI (right)	0.67	35.57 ± 1.34	73.10	72.27	0.45	42.40 ± 1.01	57.16	58.67
OBISHI	0.70	25.97 ± 1.21	**88.55**	85.23	**0.60**	**22.81 ± 0.84**	**90.92**	**77.19**
OBISHI+HIC	**0.77**	**23.97 ± 1.16**	88.21	**86.13**				

4. Evaluation

Results (Testing Phase)

- **Close-set track**

  Training/development (train): 4,863

  Evaluation (test): 2,421

- **Open-set track**

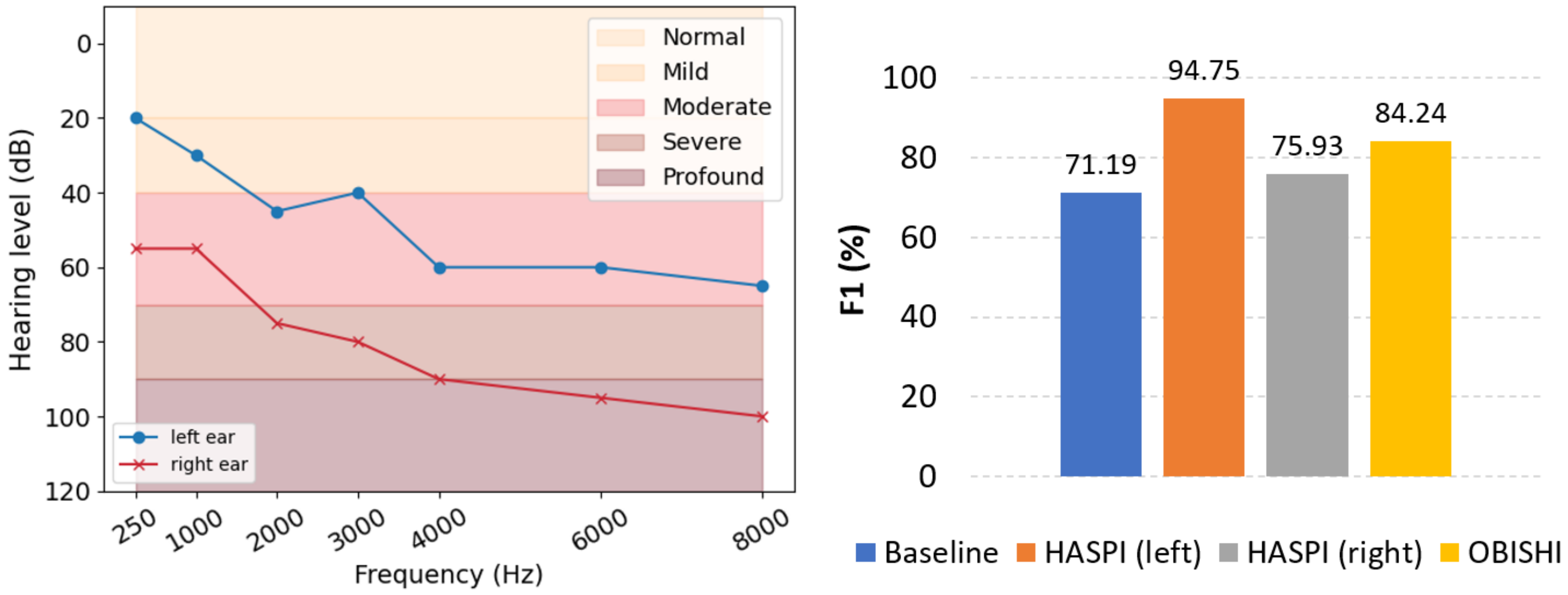
  Training/development (train-indep): 3,580

  Evaluation (test-indep): 632

Method	Track 1 (close-set)				Track 2 (open-set)			
	ρ	RMSE	F1 (%)	AUC (%)	ρ	RMSE	F1 (%)	AUC (%)
Baseline	0.62	28.52 ± 0.58	81.83	75.74	0.53	36.52 ± 1.35	68.39	68.74
HASPI (left)	0.60	37.72 ± 0.60	68.33	68.56	0.57	37.87 ± 1.20	67.88	68.58
HASPI (right)	0.60	37.66 ± 0.60	68.33	68.56	0.55	38.61 ± 1.23	67.05	67.99
OBISHI	**0.68**	**27.86 ± 0.54**	85.04	80.72	**0.67**	**28.29 ± 1.06**	**82.90**	**78.69**
OBISHI+HIC	0.41	37.19 ± 0.72	**85.16**	**87.11**				

4. Evaluation

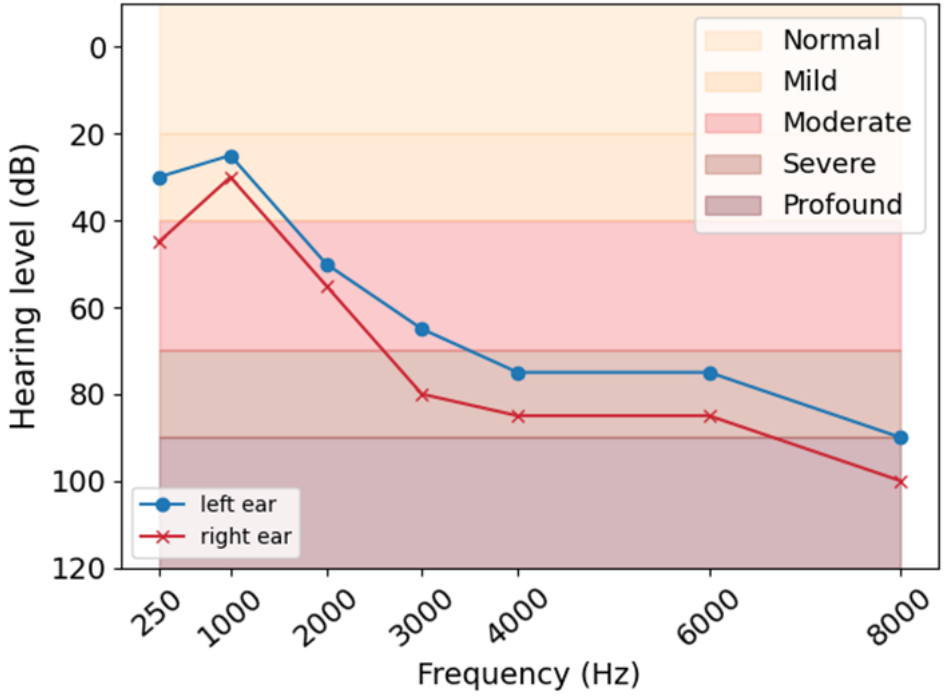
A Case Study (Listener L0217)

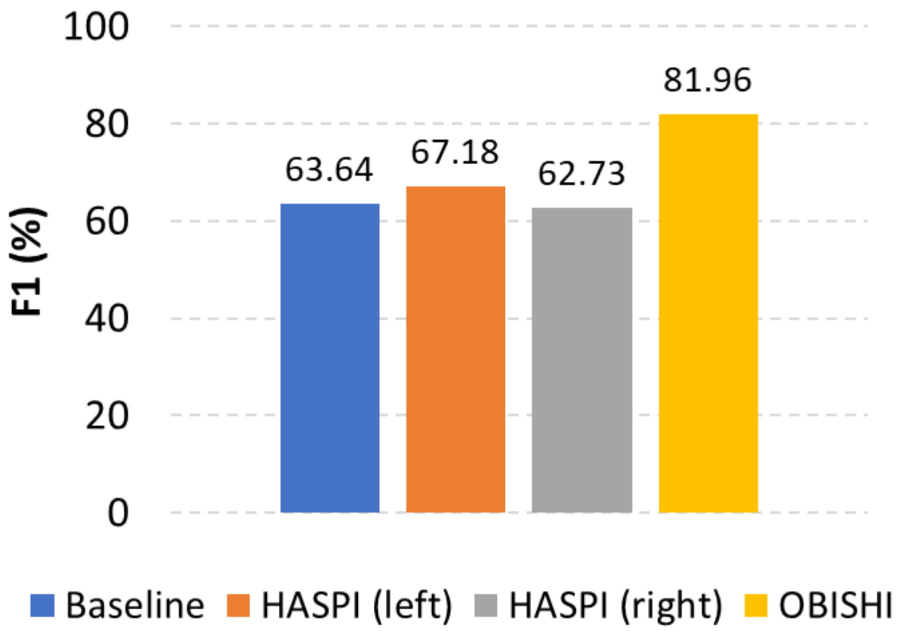


Average results of 18 evaluation scenes in total development data

4. Evaluation

A Case Study (Listener L0239)





Average results of 22 evaluation scenes in total development data

5. Conclusion

- We proposed a method, namely an objective binaural intelligibility score for the hearing impaired, OBISHI.
- The proposed intrusive metric considers multiple HI characteristics and output of a pretrained ASR system for predicting the speech intelligibility score.
- We integrated the Cambridge hearing loss model with our constructed GTFB-based predictor model in the intelligibility model.
- The evaluation was conducted with the CPC1 dataset.
- The results showed that our method without HIC predictor could significantly improve the prediction of the baseline MBSTOI and HASPI in both tracks.
- In addition, our proposed method also significantly improved the speech intelligibility prediction when the listener has a specific hearing impaired conditions in comparison to the baseline method.

THANK YOU