CPC1_EO31 System Description

Candy Olivia Mawalim, Benita Angela Titalim, Masashi Unoki
Japan Advanced Institute of Science and Technology,
1-1 Asahidai, Nomi, Ishikawa 923–1292 Japan

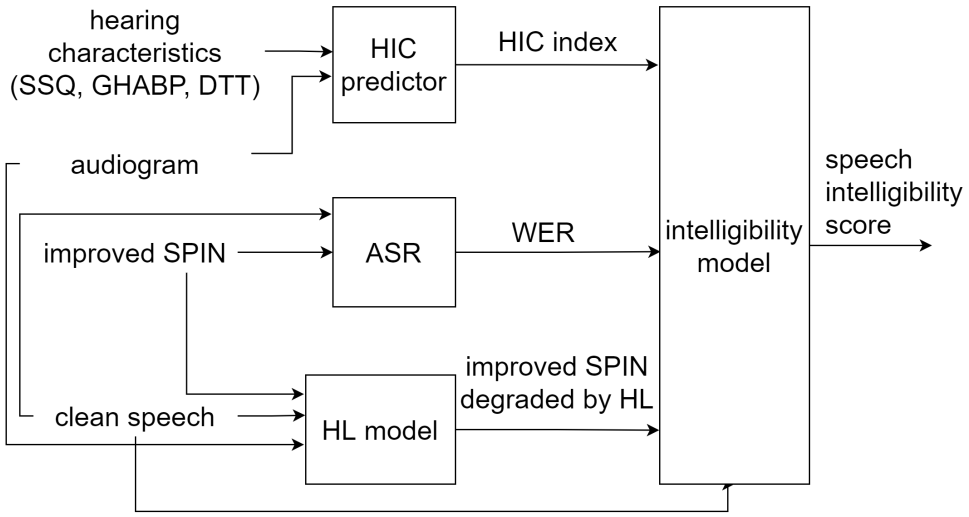

Figure 1: Block diagram of proposed method for close-set track.

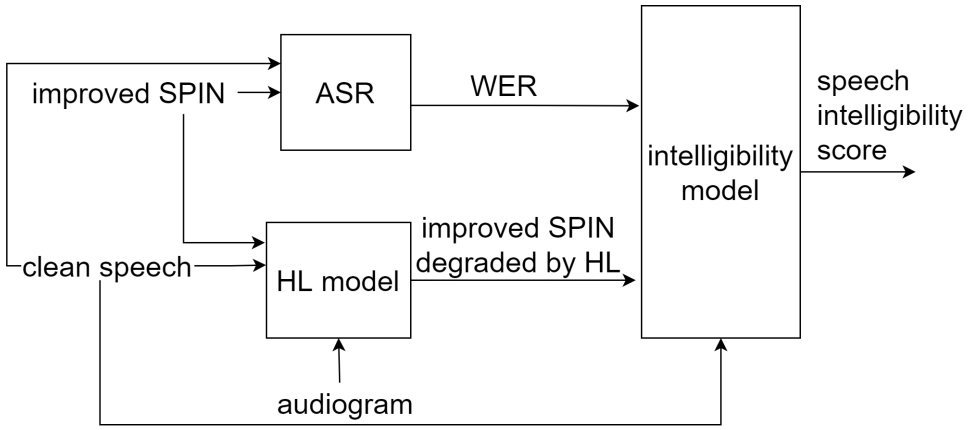

Figure 2: Block diagram of proposed method for open-set track.

This paper proposed an objective binaural intelligibility score for hearing impaired (OBISHI). This intrusive objective measurement consider the hearing impaired characteristics for predicting the speech intelligibility score. The overall process in our proposed method is shown in Fig. 1 and Fig. 2 for close-set and open-set tracks, respectively. The general inputs for both models are clean speech, improved SPIN (output of hearing aid system), and audiogram of the HI listener. The HI characteristics (HIC) of the listener, including the results of SSQ [1], GHABP [7], and DTT [2] were taken into account for inferring the HIC indices.

There are four main components in our proposed methods: HIC predictor, automatic speech recognition (ASR), hearing loss (HL) model, and intelligibility model. The HIC

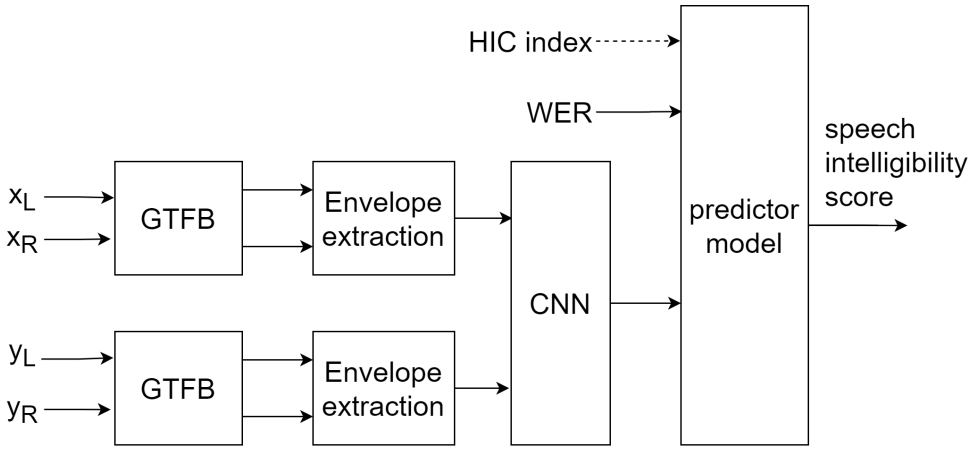

Figure 3: Block Diagram of proposed intelligibility model. x is the clean speech and y is the improved SPIN degraded by HL. The subscript letter indicates the label for ear location L for left and R for right.

predictor receives the hearing characteristics and audiogram of the listener as input and results in HIC indices as output. The HIC indices represents each characteristics. We used imputation approach to handing the missing data in HIC characteristics. The mean value is used to fill the missing data of SSQ and GHABP results from available listener. Meanwhile, we predict the DTT results using a prediction model with the input of other listener characteristics and audiogram.

The ASR receives the clean speech and the output SPIN as inputs and outputs the word error rate (WER) of the predicted sentence of the output SPIN with the predicted sentence of the clean speech as the reference. We utilized a pre-trained ASR system [6] that built using a factorized time delay neural network (TDNN-F) [5] which trained on LibiSpeech dataset [4]. The purpose of integrating an ASR in our model is to predict the difficulty of the sentence despite of HL condition (recognition rate for NH listener). The HL model developed by Cambridge Auditory Group (namely, the MSBG model) [3] was utilized to estimate the improved SPIN degradation caused by HL. The MSGB model comprises of simulations of acoustic transformation in cochlea, spectral smearing and threshold elevation, and loudness recruitment.

Figure 3 shows the block diagram of our proposed intelligibility model. We consider the speech inputs as binaural signals. An IIR time-domain gammatone filterbank (GTFB) with 32-channel was utilized to analyze the signals from both ears. Subsequently, we extracted the envelopes from the output of each channel in the GTFB analysis. These envelopes were then pass through a convolutional neural network (CNN). The final predictor model consists of two layers of fully-connected network receives the output of CNN layer, the HIC indeces, and the WER to predict the speech intelligibility score.

Bibliography

[1] Klaudia Andersson, Line Andersen, Jeppe Christensen, and Tobias Neher. Assessing Real-Life Benefit From Hearing-Aid Noise Management: SSQ12 Questionnaire Versus Ecological Momentary Assessment With Acoustic Data-Logging. *American Journal of Audiology*, 30, 12 2020.

[2] Elien Van den Borre, Sam Denys, Astrid van Wieringen, and Jan Wouters. The digit triplet test: a scoping review. *International Journal of Audiology*, 60(12):946–963, 2021.

[3] Yoshito Nejime and Brian Moore. Simulation of the effect of threshold elevation and loudness recruitment combined with reduced frequency selectivity on the intelligibility of speech in noise. *The Journal of the Acoustical Society of America*, 102:603–15, 08 1997.

[4] V. Panayotov, G. Chen, D. Povey, and S. Khudanpur. Librispeech: An ASR corpus based on public domain audio books. In *2015 IEEE ICASSP*, pages 5206–5210, 2015.

[5] Vijayaditya Peddinti, Daniel Povey, and Sanjeev Khudanpur. A time delay neural network architecture for efficient modeling of long temporal contexts. In *INTERSPEECH 2015, 16th Annual Conference of the International Speech Communication Association, Dresden, Germany, September 6-10, 2015*, pages 3214–3218, 2015.

[6] Natalia Tomashenko, Brij Mohan Lal Srivastava, Xin Wang, Emmanuel Vincent, Andreas Nautsch, Junichi Yamagishi, Nicholas Evans, Jose Patino, Jean-François Bonastre, Paul-Gauthier Noé, and Massimiliano Todisco. The VoicePrivacy 2020 Challenge evaluation plan, 2020.

[7] William Whitmer, Patrick Howell, and Michael Akeroyd. Proposed norms for the Glasgow hearing-aid benefit profile (GHABP) questionnaire. *International journal of audiology*, 53, 02 2014.