IMSE_Y4_08

Department of Computer Engineering
(Music Engineering and Multimedia)

Machine Learning for Sound Localization

Tanya KREEPANICH[1], Sutusta KUNASORN[2] and Asst. Prof. Munhum PARK[3]

Abstract

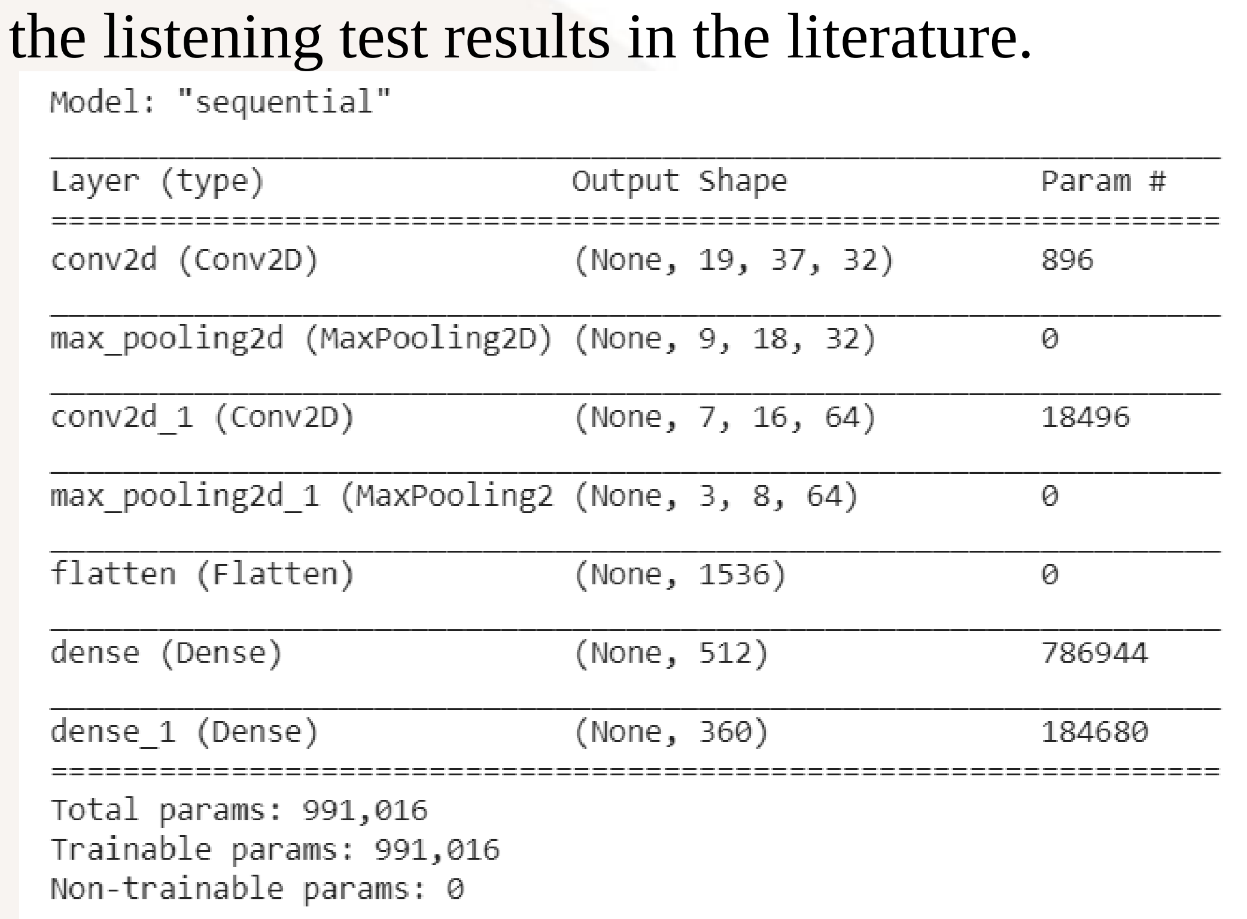
Humans can detect sound location accurately based on the difference of arrival time and sound level between two ears. This thesis is about simulating human ability of sound localization on a horizontal plane by using machine learning techniques. Our thesis is divided into 2 parts which are sound localization model and machine learning. In the sound localization model, excitation-inhibition (EI) cell activity patterns have been made to feed in the machine learning model. In the machine learning part, we have adjusted parameters for training and testing sessions to imitate human's ability and comparing the mean error and standard deviation of the predicted azimuth angle for each target angle. From our results, we found that the model structure, the number of epochs, and the properties of training and testing data influenced the model accuracy.

Introduction

Humans can accurately detect the location of sound sources, for which the interaural time difference (ITD) and interaural level difference (ILD) are the important cues in the horizontal plane [1]. The Computational models of sound usually consist of peripheral, binaural and central processes. In our study, EI patterns were created in the binaural process which contain ITD and ILD information, and passed to the central process, where we will apply machine learning (ML) to determine the azimuth angle. The objective of the study is to find the best ML model that can imitate human performance by considering its mean error and standard deviation of the prediction.

Methodology

First, we generated EI patterns to make an input for the training and testing of our ML model: 250 sets of EI patterns for training, 50 for validation, and 100 for testing. Each set consists of one EI pattern at every angle from 0 to 359 degrees. We used a main model as shown in Figure 1, based on which we tried different. Then, we analyzed model predictions in terms of the mean error and the standard deviation and compared them with the reference data, the mean error and the standard deviation of the listening test results in the literature.

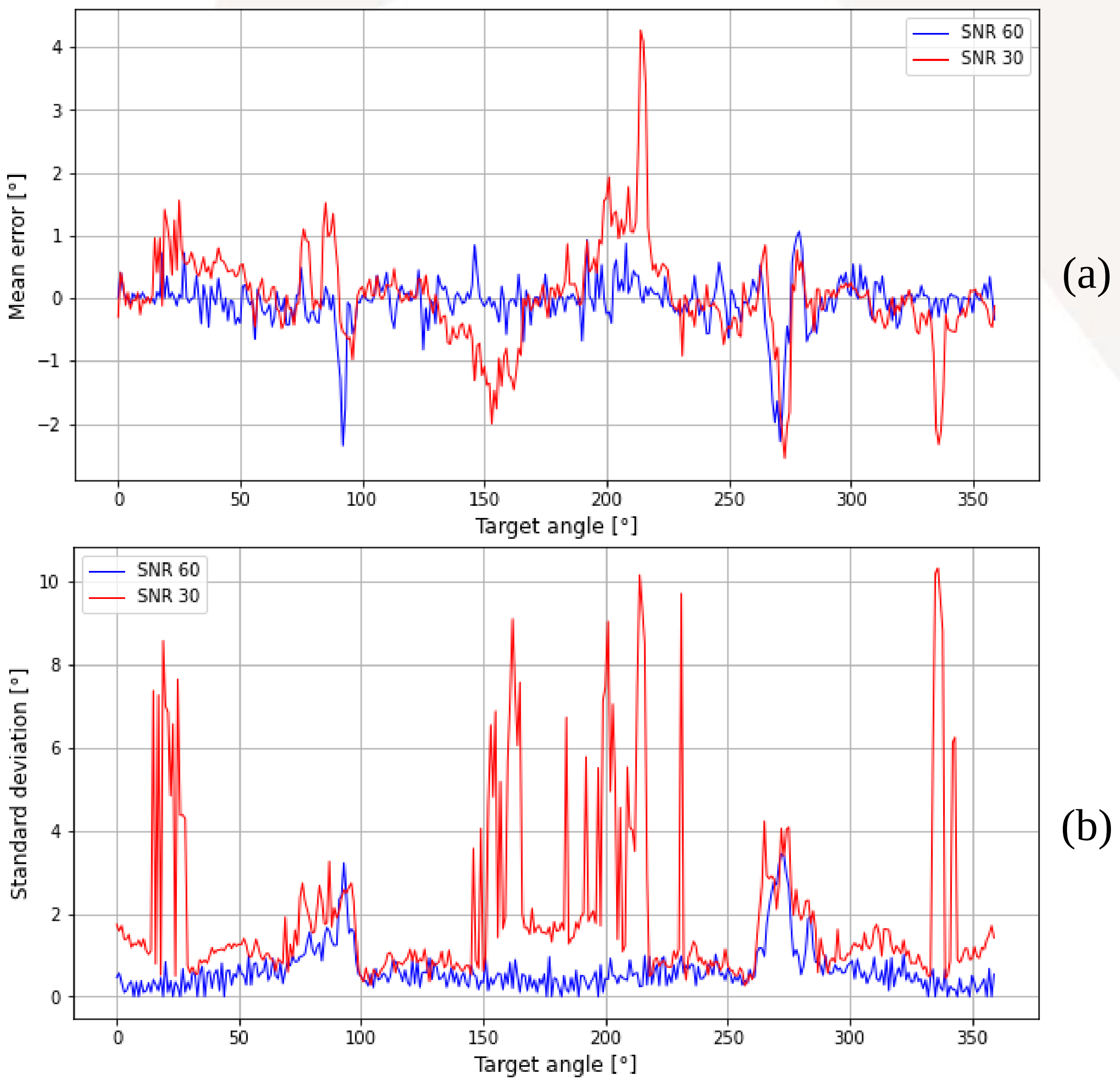


```
Model: "sequential"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
conv2d (Conv2D)              (None, 19, 37, 32)        896
_________________________________________________________________
max_pooling2d (MaxPooling2D) (None, 9, 18, 32)         0
_________________________________________________________________
conv2d_1 (Conv2D)            (None, 7, 16, 64)         18496
_________________________________________________________________
max_pooling2d_1 (MaxPooling2 (None, 3, 8, 64)          0
_________________________________________________________________
flatten (Flatten)            (None, 1536)              0
_________________________________________________________________
dense (Dense)                (None, 512)               786944
_________________________________________________________________
dense_1 (Dense)              (None, 360)               184680
=================================================================
Total params: 991,016
Trainable params: 991,016
Non-trainable params: 0
```

[Figure 1 Structure of the main model]

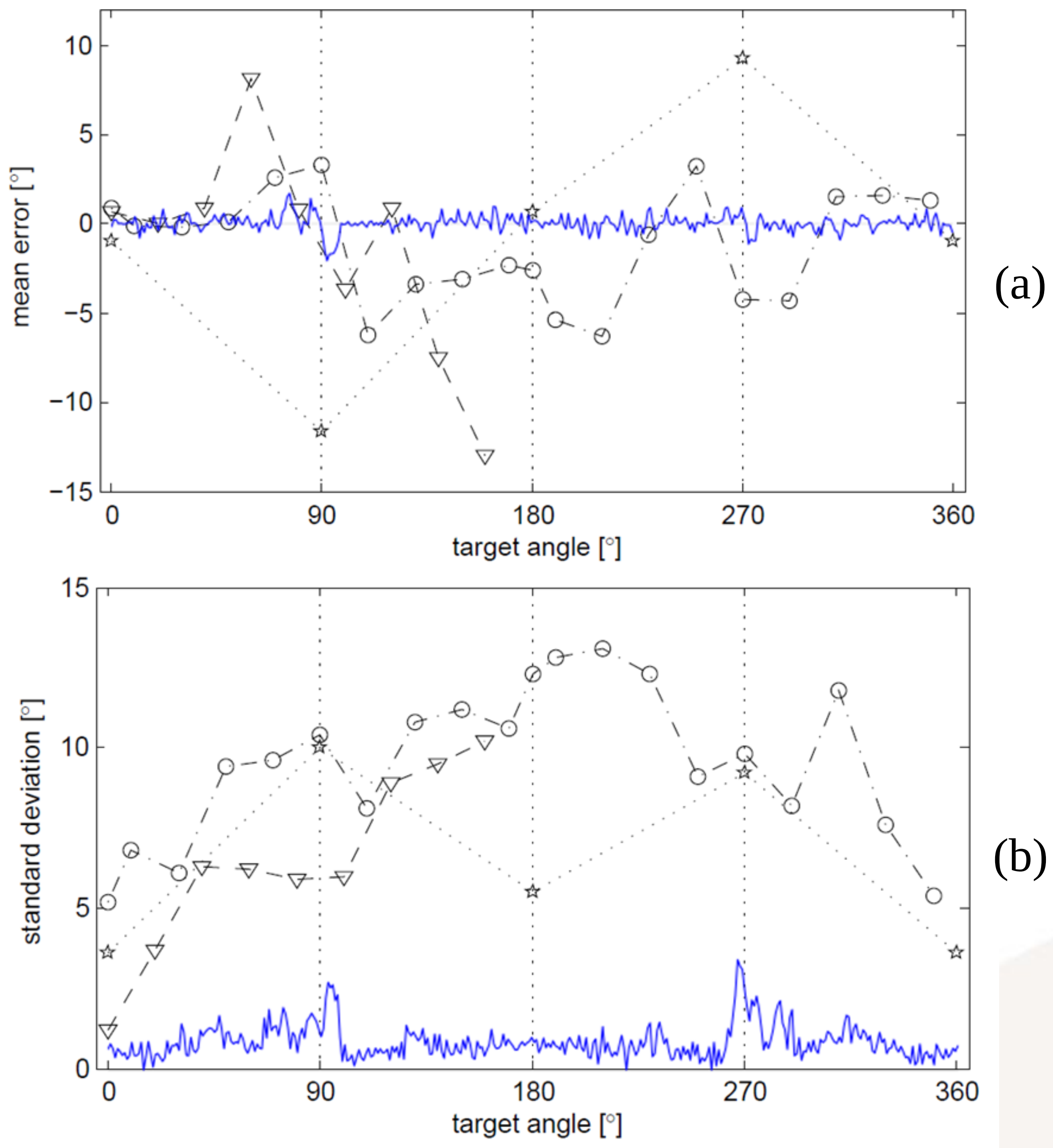
Results

The results of the comparison between different SNR values of the training data (see Figure 2) are as expected. In Figure 2 (a), we can see that the mean error of the model trained with SNR=30 dataset is quite higher than the model trained with SNR=60. It is interesting that the model which is trained and tested by SNR=30 dataset is having lower accuracy than the model which is trained and tested by SNR=60 and SNR=30 dataset, respectively. Since SNR means signal-to-noise ratio, it means the higher SNR is, the less ambient noises are. Therefore, EI patterns created at a high SNR can contain clean features at each azimuth angle, improving the model performance. This is why we prefer to use the model trained at a higher SNR value dataset.



[Figure 2 (a) Mean error and (b) standard deviation comparison between a model which is trained by EI patterns with SNR=60 (blue) and EI patterns with SNR=30 (red).]

In figure 3, we can see the performance of the best model from the experiment in comparison with some of the published listening test results. In terms of the mean error, the results from our experiment did not resemble those from the listening test. Nevertheless, the following observations may be made: For the target angles from 0 to 50 degrees, the mean errors in our experiment are in the range of ±1 degree, quite like the listening test published by Carlile et al. and Makous and Middlebrooks. In terms of the standard deviation, the values from our experiment are significantly high at around 90 and 270 degrees, similar to the listening test data published by Blauert and Carlile et al. (see Park [1] for the reproduced data).



[Figure 3 Comparisons of (a) the mean error and (b) the standard deviation between our model and some of the published listening test results: Blauert (★), Carlile et al. (o), Makous and Middlebrooks (▽) [1] and our model (blue).]

Conclusion

In this study, machine learning has been applied to imitate human ability of sound localization in the horizontal plane. We generated EI patterns as the input to the model and analyzed the model predictions in terms of the mean error and the standard deviation. We carried out various experiments to see which model gave results similar to the human performance. We found that the model without any convnet layers is best, which was trained with high-SNR EI pattern (SNR=60) for less than 20 epochs.

References

[1] Park, M. (2007). "Models of binaural hearing for sound lateralisation and localisation", University of Southampton, Institute of Sound and Vibration Research, PhD Thesis, 125-128.

E-mail: 60010453@kmitl.ac.th[1], 60011089@kmitl.ac.th[2] , munhum.pa@kmitl.ac.th[3]