On a Binaural Model with Front-back Discriminator using Artificial Neural Network trained by multiple HRTF catalogs

Shun Yoshino[1]; Takuro Tomita[2]; Yoshifumi Chisaki[3]; Tsuyoshi Usagawa[4]

[1,2,3,4] Kumamoto University, Japan

ABSTRACT

Various binaural models have been proposed for the application of hearing assistance system as well as humanoid robot, and a frequency domain binaural model(FDBM) is the one. Like other binaural models, the original FDBM can separate and segregate a signal from the specific direction based on interaural information, but it works only in the frontal semisphere due to front-back confusion. In order to reduce this confusion, a front-back discriminator was proposed for FDBM using artificial neural network (ANN) trained by a head related transfer function (HRTF) catalog. This discriminator has strong dependency on the trained HRTF catalog thus it is not robust against various fluctuation including individual differences and reverberation. This paper discusses an extention of the discriminator using multiple HRTF catalogs for ANN training. The simulation results for the new discriminator show the possibility to reduce the front-back confusion under various conditions including ones obtained in a reverberant room.

Keywords: Sound source direction, Head Related Transfer Function(HRTF), Frequency Domain Binaural Model(FDBM), Front-back confusion
I-INCE Classification of Subjects Number(s): 74.6

1. INTRODUCTION

There are many proposals for an estimation of sound source direction and a separation of a specific sound source using a microphone array. Human being also has a capability to detect the source's direction as well as to segregate the focused sound based on a binaural hearing; two ears with the effect of head and torso. The binaural model is understood as a computer model to simulate binaural hearing capability of human being(1).

A Frequency Domain Binaural Model (FDBM)(2) is one of the binaural models to segregate a specified sound source. FDBM has two inputs corresponding to left and right ears based on the Interaural Level and Phase Differences (ILD and IPD) for each frequency bin, and two outputs which maintain the binaural cues for segregated signal components. Although FDBM works well for frontal side, it has a corn shape ambiguity, so-called "corn of confusion" in the field of psychoacoustics. In the previous paper, a FDBM with front-back discriminator was proposed to extend the conventional FDBM. This discriminator was realized by means an Artificial Neural Network (ANN) which trained by spectral cues of Head Related Transfer Function (HRTF) catalog(3).

This discriminator works well under the same condition of ANN training, however, it has a strong dependency on individual characteristics of HRTF catalogs.

Even if the discriminator is built with a HRTF catalog for the specific Head and Torso Simulator (HATS), the error of front-back discriminator increased for other instance of the same model of HATS. This issue may come from the over-trained ANN to the specific characteristics of HRTF catalog or from the individual difference between HATSs, or fluctuation of measurement environment.

In this paper, an effect of HRTF catalog dependency on the FDBM is examined. In addition, in order to reduce individual difference of dummy head and fluctuation of measurement environment, a method of front-back discrimination using ANN trained by multiple HRTF catalogs is proposed. The simulation is carried out to confirm the possibility to reduce the front-back confusion using the extended discriminator. Also an example result under an actual reverberant condition is also discussed.

[1] shun@hicc.cs.kumamoto-u.ac.jp
[2] takuro@hicc.cs.kumamoto-u.ac.jp
[3] chisaki@hicc.cs.kumamoto-u.ac.jp
[4] tuie@cs.kumamoto-u.ac.jp

2. FDBM WITH FRONT-BACK DISCRIMINATOR

The original FDBM utilizes interaural information, ILD and IPD, for each frequency bin after Fast Fourier-Transformation (FFT) as shown in Fig.1. Observed IPDs and ILDs are compared with stored IPD and ILD information in the database and the frequency components corresponding to the specified directional range are selected by Segregation Filter shown lower part of this figure. Because of the quasi-symmetrical shape of the head and torso, IPD and ILD information for each frequency bin is not sufficient to determine whether the sound comes from frontal side or back side. This means the front-back confusion hardly solved based on the information of individual frequency bin.

On the other hand, psychoacoustical studies show the evidences that human beings have a capability to solve “cone of confusion” (4) which makes front-back confusion. Also Iida *et al.*(5) show that the spectral cues contribute to solve the front-back confusion and there are systematic relationship between notches of HRTF gain characteristics and rising angle of sound source in the sagittal coordindate shown in Fig.2. As shown in this figure, the sagital coordinate is defined by two parameters; a lateral angle α and a rising angle β. Iida *et al.* show that a lateral angle relates to IPD and ILD while spectral cues across the frequency range related to a vertical angle. Especially, in the frequency range above 6kHz, notches in HRTF gain characteristics systematically related to the rising angle in sagittal coordinate. Furthermore, first notch and second notch are dominate spectral cues of rising angles perception.

Based on this evidence, the front-back discriminator is implemented on ANN to figure out the relative relationship across the frequency range and inserted between FFT and segregation filter in FDBM as shown in Fig.1.

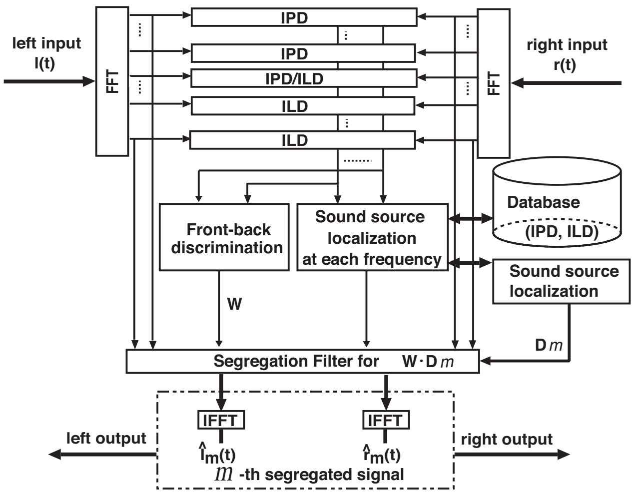


Figure 1: A block diagram of Frequency Domain Binaural Model with front-back discriminator.

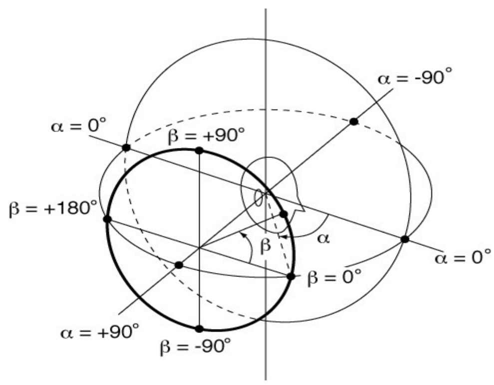


Figure 2: The sagittal coordinate system. A lateral angle : $-90^\circ \leq \alpha \leq +90^\circ$, and a vertical angle :$-180^\circ \leq \beta \leq +180^\circ$.

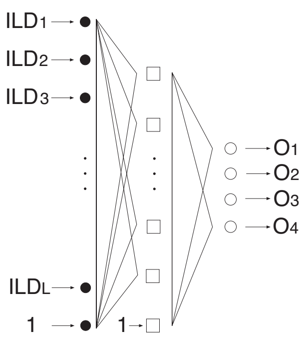


Figure 3: Structure of ANN. (•: Neurons of input layer,□: interlayer, ○: output layer.)

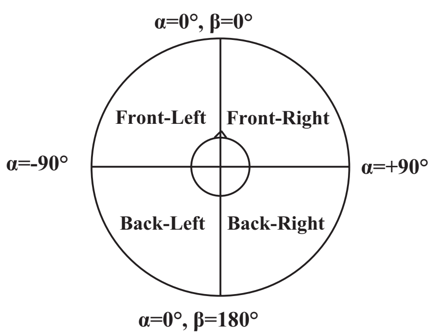


Figure 4: Quadrant segmantation.

The front-back discriminator is implemented on ANN as shown in Fig.3 which has L ILDs across the frequency range and one fixed input to standardlize the input level. Also it has 4 outputs corresponding to quadratic areas (Front-Left(FL), Back-Left(BL), Front-Right(FR), Back-Right(BR))(3). The segregation filter in Fig.1 selects the frequency components which come from the specific range of α as well as quadrant shown in Fig.4.

3. DEPENDENCY OF FRONT-BACK DISCRIMINATOR ON HRTF CATALOG

3.1 Bakground of discriminator design

The performance of front-back discriminator depends on the configuration of ANN including selection of input signal, the number of inputs and outputs, the number of layers, as well as training process. Previous works provide that the current configuration of ANN shown in Fig.3 can estimate the direction of sound sources in almost whole spherical areas by (α, β) under high SNR condition(6). Also ILD has an advantage to reduce the effect of spectral information of sound source and it is also very convenient to compare with ILD information already stored in FDBM. Based on the various trials in the former works, the four outputs corresponding to the quadratic segmentation of sphere provide the stable and highest performance among examined. Based on those discussion, the training dataset for ANN is focused discussed in this paper.

3.2 Training process of ANN

In order to discuss the dependency of front-back discriminator on HRTF catalogs, two sets of HRTF catalogs, obtained from B&K 4128 Head and Torso Simulator (HATS) and KEMAR dummy head(7).

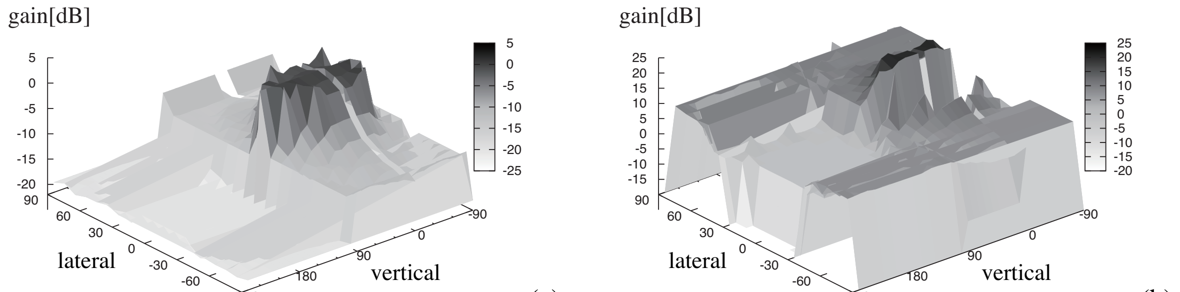
Table 1 shows common basic parameters for ANN training data for the FDBM. In training process, almost all spherical HRTF information in which the directions corresponding to own torso and body are excluded, is used.

Table 1: Condition of ANN training.

HRTF catalog : KEMAR and B&K	
Sampling rate	16 kHz
Quantization bit	16 bit
Frame length	512
Number of Frame	20
Frequency resolution	31.25 Hz
Sound sources	White noise

Table 2: Comparison of the estimation accuracy on front-back discrimination.

Estimation accuracy [%]	Input signal HRTF	
	KEMAR	B&K
KEMAR ANN	99.7	65.1
B&K ANN	47.0	99.5



(a) KEMAR ANN (KEMAR signal). (b) KEMAR ANN (B&K signal).

Figure 5: Directivity pattern obtained using the front-back discriminator trained by KEMAR HRTF catalog. The vertical axis shows normalized gain to the gain at $(0^\circ, 0^\circ)$ and where -25 dB is used for flooring.

3.3 Simulation on Catalog Dependency

Table 2 shows simulation results obtained by two types of front-back discriminators; one is trained by B&K 4128 HATS (labeled B&K) and the other one is by KEMAR dummy head (labeled KEMAR) against two sets of input signals, obtained by B&K HRTF and KEMAR HRTF, are examined. The results show the accuracy of quadratic segmentation in horizontal plane, $\beta = 0^\circ$ and $-90^\circ \leq \alpha \leq 90^\circ$ in every 10°.

As shown in this table, the accuracies are higher than 99% when the training set and test set are the same HRTF catalog. However accuracies drastically drop when the HRTF catalogs are the different such as 47% when KEMAR data fed into B&K based ANN.

Figure 5 shows the directional characteristics as the relative gain for whole range of sphere when the the segregation range is set to $-30° \leq \alpha \leq 30°$ of the frontal side; $-90° \leq \beta \leq 90°$. As shown in Fig. (a), the discriminator trained by KEMAR catalog works well for the inputs generated by KEMAR. However it does not work well for the input generated by B&K catalog as shown in Fig. (b). This figure also show the dependency of HRTF catalog for segregation task using a front-back discriminator. Also Fig. (b) shows the front-back confusion mainly in large lateral angle range, and this phenomenon is easy to understand because it is not easy to distinguish front and back when the sound source is far left or far right direction.

4. A PROPOSED TRAINING METHOD OF DESCRIMINATOR

As shown in the previous section, there is a strong dependency on HRTF catalogs used for training and discrimination. This dependency is also obtained even if the discriminator uses for B&K HATS under real-time experiment using B&K HRTF catalog provided by other institute. The degradation due to different instances of the dummy head comes from various reason including individual difference, difference of measurement condition, as well as other kind of interferences.

In this paper, the authors challenge to reduce the dependency on individual differences in a dummy head by means of addition supplemental HRTF sets for the training stage of the front-back discriminator.

As the fundamental HRTF catalog, the full set HRTF catalog provided by Nagoya University(8) is used. Using the same B&K 4128 HATS, but different individual peace, the authors measure 5 sets of HRTF for every 10° in horizontal plane; $-90° \leq \alpha \leq 90°$ and $\beta = 0°$.

In the following section, the simulation and real-time experiment are conducted using some sets of measured HRTF combined with provided HRTF catalog. In order to distinguish each condition, the following short expressions are used: **Full set** means HRTF catalog for sphere provided by Nagoya University, **Full set and supplement(s)** means the combination of full set HRTF catalog with newly measured horizontal HRTF set.

5. RESULTS OF SIMULATION AND REAL-TIME EXPERIMENT

5.1 Simuation

Five sets of measured HRTF on horizontal plane are used in the simulation combined with Full set HRTF catalog.

Closed Test :

Four sets of measured HRTF are used as the supplmental catalog to Full set for the training the front-back discriminator, namely ANN trained by Full set and 4 supplements. Test of discrimination is performed using the one of 4 measured HRTF sets for generating the simulation input signals.

Open Test :

Training of ANN is performance using Full set and 4 supplements, however, the test is performed using the remaining HRTF set which is not used for training.

Table 3 shows the results the estimation accuracy of the front-back discriminators under various conditions. Two ANNs are examined ; one is trained only under Full set condition, and the other is trained under Full et with 4 supplements as described in the previous section. The asterisks mark the results under open test condition while ones without asterisk show closed test. According the results shown in this table, the improvement by means of supplemental HRTF sets for training is 12.5% (from 61.9% to 74.4%) for closed test condition using Full set input, and 13.7% (from 66.3% to 80.0%) for open test condition using Supllement HRTF input.

There is 4.2% (61.9% to 66.1%)improvement from Full set input to Supplement input using Full Set ANN. It is not yet very clear why the improvement is obtained, however, the authors interpreted as the effect of amount of training data for horizontal plane because supplemental sets have only HRTF on horizontal plane.

Table 3: Comparison of the estimation accuracy on front-back discrimination. In the table * represents the results of “open test”.

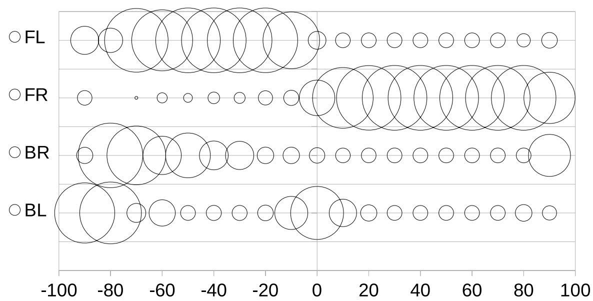
	Estimation Accuracy [%] for Input Signal generated by		
HRTF set used for Training	Full Set	Supplement	Supplement
Full set	61.9	66.1 *	66.3 *
Full set with 4 supplements	74.4	83.1	80.0 *

Figure 6 shows the details of averaged ANN outputs; FL (Front Left), FR (Front Right), Back Right (BR) and Back Left (BL). The horizontal axis shows the lateral angle of sound source, and open circles shows the ratios for each 4 outputs; the total of 4 outputs is always 100%. Plots (a) and (c) represent the results obtained by Full set ANN and 4 Supplements ANN, respectively. Also plots (b) and (d) correspond them as well. Plots (a) and (b) are results obtained when the sound source is in frontal side, $\beta = 0°$, while plots (c) and (d) are ones when source is in back side, $\beta = 180°$.

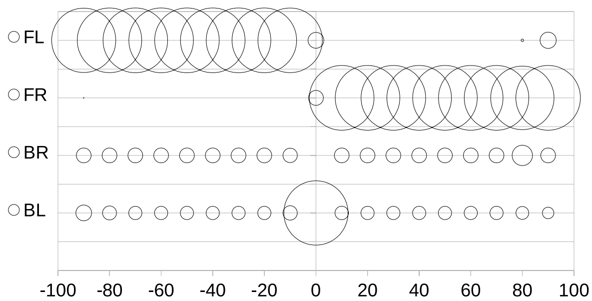
In comparison with plots (a) and (b), the estimated quadrants for frontal side sound source are concentrated into frontal side for plot (b), thus the training by supplemental HRTF provides improvement of discrimination performance.

On the contrary, plots (c) and (d) show some extent of improvements by reducing the occurrence of front-back confusion as shown in plot (d), there are some front back errors in the range of large lateral angles, especially on right hand side.

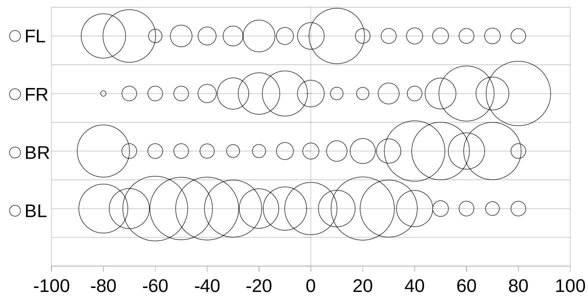
As the results, the supplemental sets for training ANN is effective to improve the performance of front-back discriminator in some extent. However further improvement is necessary.



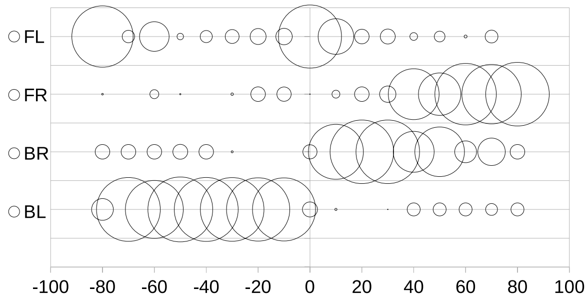
(a) ANN trained full set catalog (front).



(b) ANN trained full set catalog + 4 supplement (front).



(c) ANN trained full set catalog (back).



(d) ANN trained full set catalog + 4 supplement (back).

Figure 6: The front-back discrimination accuracy of FDBM implemented front-back discriminator in horizontal plane. Input signal is subset HRTF which is not used training of all ANN.

5.2 Result of Real-Time Experiment

Real-time operation of front-back discriminator on PC-based FDBM implementation is conducted within an ordinary laboratory room as shown in Fig. 7. Experiment is conducted using B&K 4128 HATS on the turn table with a loudspeaker on the stands whose height is leveled to the ear position of B&K HATS. Direction of sound is controlled by rotating the HATS instead of changing the loudspeaker's position. The performance of front-back discriminator is measured as the correct discrimination ratio for 3000 trials for every 10°.

Figure 8 provides the correct discrimination ratio for direction. The solid thin circle line corresponding 100%. Black solid thick line shows the performance obtained by Full set ANN while red broken thick line shows one obtained ANN trained by Full set and 4 Supplements HRTF. The ideal plot is the 100% circle, thus the performance obtained by Full set and 4 Supplements ANN is good in Front-Right and Back-Left, however it is degraded largely in Front-Left and Back-Right. The reason of degradation is not yet clear and there are lots of challenging issues to make a front-back discriminator robust enough for actual acoustic condition.

6. CONCLUSION

In this paper, trials to make a front-back discriminator embedded in Frequency Domain Binaural Model are shown. Simulation results show the potential to improve the discrimination performance by adding supplemental HRTF catalog into full set HRTF catalog for training ANN used for discriminator. However the improvement is limited under simulation and the experimental result under real-time condition tells us

Figure 7: Real-time Experimantal Setup.

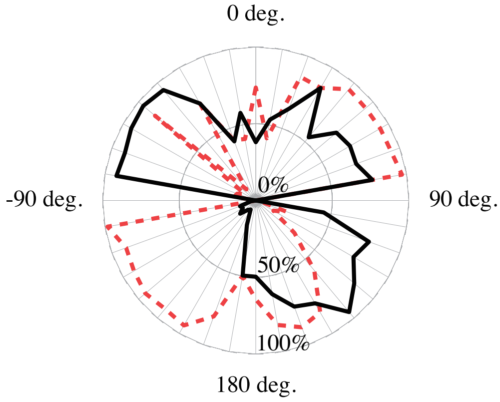


Figure 8: Discrimination peformance for real-time experiment.

necessity of further improvement.

REFERENCES

1. Blauert J. Spatial Hearing - Revised Edition. MIT Press; 1996.

2. Nakashima H, Chisaki Y, Usagawa T, Ebata M. Frequency domain binaural model based on interaural phase and level difference. Acoustical Science and Technology. 2003;24(4):172–178.

3. Saho A, Imamura K, Chisaki Y, Usagawa T. Improvement of directional selectivity of binaural hearing assistance system- On front-back confusion with an artificial neural network -. Proc Kyushu-Youngnam Joint Conference on Acoustics. 2011;p. 81–84.

4. Morimoto M, Aokata H. Localization cues of sound sources in the upper hemisphere. Journal of the Acoustical Society of Japan. 1984;5(3):165–173.

5. Iida K, Itoh M, Itagaki A, Morimoto M. Median plane localization using a parametric model of the head-related transfer function based on spectral cues. Applied Acoustics. 2007;68:835–850.

6. Usagawa T, Saho A, Chisaki Y. A binaural hearing assistance system with front-back discrimination capability. Proc Acoustics. 2012;p. 1–6(CD–ROM).

7. Nakado T, Nishino T, Takeda K. Head-Related Transfer Function measurement in sagittal and frontal coordinates. Journal of the Acoustical Society of Japan. 2008;29(5):335–337.

8. "Acoustics Speech and Behavior Signal Processing Laboratory at Nagoya University", "Itakura Laboratory", "Center for Integrated Acoustgc Information on Research". Head Related Transfer Function (HRTF) Database;. Available from: `http://www.sp.m.is.nagoya-u.ac.jp/HRTF/`.