Front. Electr. Electron. Eng. China (2006) 4: 455–459
DOI 10.1007/s11460-006-0087-z

RESEARCH ARTICLE

ZHANG Jie, MA Hao, WU Zhen-yang

Optimal approximation of head-related transfer function's pole-zero model based on genetic algorithm

Abstract In the research on spatial hearing and virtual auditory space, it is important to effectively model the head-related transfer functions (HRTFs). Based on the analysis of the HRTFs' spectrum and some perspectives of psychoacoustics, this paper applied multiple demes' parallel and real-valued coding genetic algorithm (GA) to approximate the HRTFs' zero-pole model. Using the logarithmic magnitude's error criterion for the human auditory sense, the results show that the performance of the GA is on the average 39 % better than that of the traditional Prony method, and 46 % better than that of the Yule-Walker algorithm.

Keywords optimal approximation, HRTFs, pole-zero model, genetic algorithm

1 Introduction

It is a well-known fact that the human auditory system works well at determining a sound's positions in an acoustic environment. People have observed this phenomenon long ago, and have tried to investigate and interpret it. So far, there are many mathematical models aiming to solve this problem. The most classic and effective solution is the duplex theory, which utilizes the interaural time difference (ITD) and interaural intensity difference (IID) to interpret the cause of spatial hearing. Because of these results, there have been some tentative applications in the auditory display of multimedia systems, for instance VoiceNotes of MIT Media Lab. However, there are still no completely successful models capable of explaining some of the phenomena associated with spatial hearing, such as the perception of sound's position at an elevation, the "cone of confusion", and so on.

In the succeeding research, people take cognizance of the cues implied in grotesque, but a little inerratic spectrum. Thus, HRTF has been introduced to describe the entire information related to spatial hearing, where ITD can be seen as the delay between two ears and IID can be the difference of HRTFs' magnitude. People have constructed different mathematical models to simulate this delicate function, and have tried to give a felicitous interpretation for this phenomenon [1, 2].

As a popular model structure, the HRTFs' pole-zero model has been greatly studied for this field [3–6]. Basically, as for the design of HRTFs' pole-zero model used in VAS application, there are two conventional methods, the Prony and Yule-Walker, which can obtain proper results. Unfortunately, these methods only take into account the self pertinence of the HRIR's response sequence, and ignore the particularity of the different frequency of a human being's spatial hearing. Reference [5, 6] went into the particulars of this problem. The results have shown that, based on the log-least square error criterion, the simple, binary-coded GA used in Ref. [5] excels in the steepest descent method which was brought forward in Ref. [6], especially in the efficiency of computation. However, as for the comparison with the conventional methods, such as Prony, Yule-Walker, and so on, GA's superiority has not been given prominence. In our work, we think of the similarity of HRTFs, and apply a multi-group, real-valued GA to approximate the pole-zero model's design. Simulation results show that the improved GA is obviously superior to the traditional methods.

Synthetically, the paper is organized as follows. In Sect. 2, the data of HRTFs used for simulation are described. The improved GA's procedures with the simulation results are presented in detail in Sect. 3. After that, a conclusion is given in Sect. 4 including several steps in the direction of

Translated from *Journal of Southeast University* (*Science and Technology*), 2006, 36(1): 19–24 (in Chinese)

ZHANG Jie (✉)
The CETC 14th Research Institute, Nanjing 210013, China
E-mail: zjhf1978@yahoo.com.cn

ZHANG Jie, MA Hao, WU Zhen-yang
College of Information Science and Engineering,
Southeast University, Nanjing 210096, China

future study.

2 HRTFs data for simulation

Here, the data of KEMAR's HRTFs, which was offered by MIT Media Lab [7], are used in our simulation work. The measurements of this data package were made with speakers on discrete positions at every 10° of elevation, and at 5°–30° unequally in the azimuth. Table 1 gives the list of detailed sample points.

Table 1 Measurement of KEMAR's HRTFs

Elevation/°	Number	Step/°
–20–20	72	5.00
–30–30	60	6.00
–40–40	56	6.43
50	45	8.00
60	36	10.00
70	24	15.00
80	12	30.00
90	1	—

Moreover, the measured data may be contaminated by deficient factors. Thus, it is necessary to get rid of the contamination before further processing. Some meticulous and important processing refers to Ref. [7]. Before optimizing the model's parameters, the original HRTFs are equalized to get the CTF and DTF, which represent the common component for all positions and the own components related to different positions, respectively. In the next process, the algorithms are also compared to approximate the DTFs.

3 GA's application on IIR model of DTFs

GA is a stochastic search procedure modeled on the Darwinian concepts of natural selection and evolution. In GA, a set of potential solutions is allowed to evolve towards an optimal solution. Evolution toward an optimal solution occurs as a result of the pressure exerted by a fitness-weighted selection process and the exploration of the solution space. This is accomplished by recombination and the mutation of existing characteristics present in the current population. The GA is particularly effective when the goal is to find an approximate maximum in a high-dimension, a multimodal function domain in a near-optimal manner. The ability of the GA to perform in difficult optimization problems is further enhanced when the problem can be cast in a combinatorial form. In our study of spatial hearing, the importance of DTFs on different frequencies can just be combined into the optimization process by a proper fitness-weighted function's design to get better results.

In the following, detailed procedures and choices of the GA's parameters are presented to improve the algorithm's efficiency and guarantee an optimum solution. Thereunto, we utilize the GA toolbox supplied by the University of Sheffield [8], and we sincerely express our thanks.

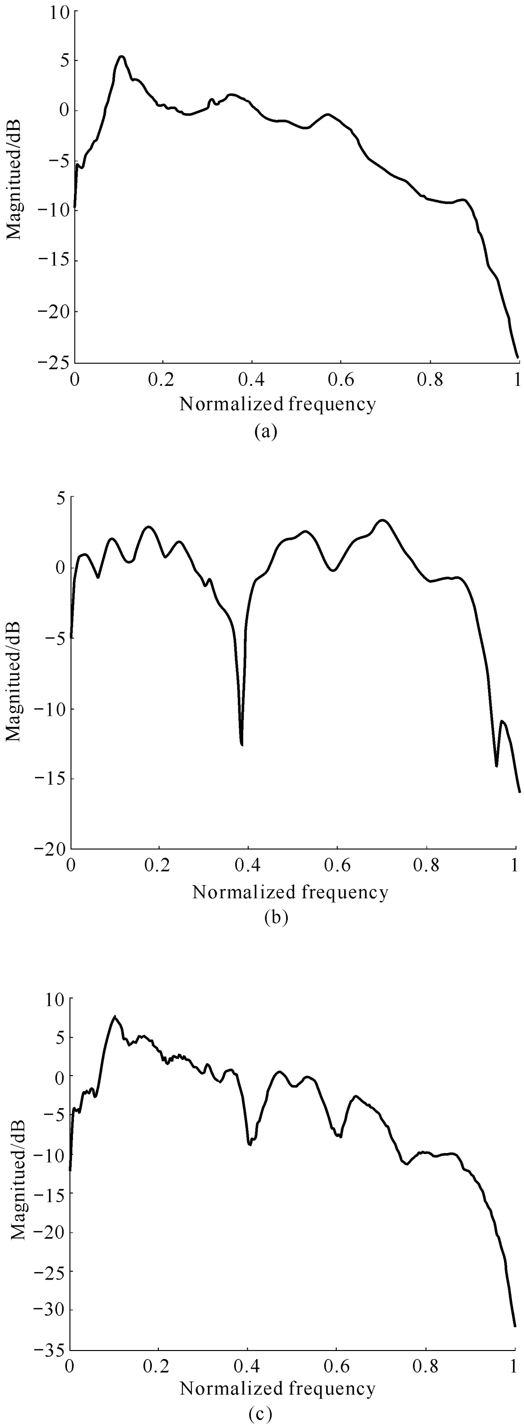

Fig.1 A demonstration of the HRTFs' equalization (elevation = 0°, azimuth = 30°). (a) CTF; (b) DTF; (c) HRTF

1) Multiple demes' parallel genetic algorithm As for algorithmic structure, in view of the randomness of the GA and the multi-local peaks of the solution space, we apply multiple demes' parallel evolution to divide the original group into several subgroups, which efficiently retains evolutional stability and avoid a premature convergence of the GA. The migration strategy among the individual subgroups is unrestricted migration as shown in Fig. 2 [8].

Here, 6 subgroups and 20 individuals in each group are set in our program. The results show that the multiple demes' parallel configuration excel the single group for the search of an optimal solution.

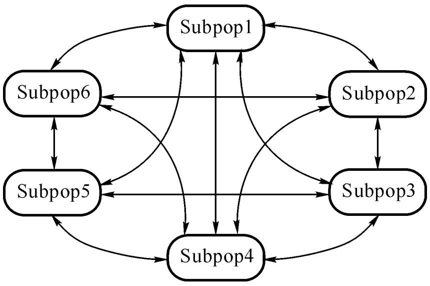

Fig. 2 Unrestricted migration topology among sub-group

2) Encoding and decoding We use the GA to approximate the pole-zero model of the DTF, where the model's parameters and structure are given below:

$$F:\{\{b_1, b_2, \cdots b_M\}, \{a_1, a_2, \cdots, a_N\}\}$$

$$H(\theta, z) = \frac{\sum_{k=0}^{M} b_k(\theta) z^{-k}}{1+\sum_{k=1}^{N} b_k(\theta) z^{-k}} = \frac{B(\theta, z)}{A(\theta, z)} \tag{1}$$

Equally,

$$H(\theta, z) = \frac{B(\theta, z)}{A(\theta, z)} = \frac{C_z - Q_1 \prod_{i=1}^{Q_2} \left[1 - q_i(\theta) z^{-1}\right]}{\prod_{i=1}^{P} \left[1 - p_i(\theta) z^{-1}\right]} \tag{2}$$

where b_k (or q_i) and a_k (or p_i) are the numerator and denominator of the pole-zero model; and the constants in the numerator and denominator are all set to 1. The aim of our work is to search for optimal parameters in the solution space F under the model Eqs. (1) and (2) of $H(\theta, z)$.

In view of the minimum phase characteristics related to the system's magnitude, we can encode the model's parameters in a unit circle to limit the bound of the original individual for an efficient search in the solution space. As for simple GA, the parameters are encoded in binary. But for complex and high precision problems, the binary code is not convenient to reflect the problem's characters and the performance is usually poor. Here, we use real-valued coding of the pole-zero's parameters in our work; and at the same time, under the circumstances of a real-valued numerator and denominator, the positions of the poles and zeros symmetrically conjugate and have superposition's characteristic. Figure 3 shows an example of the original individual.

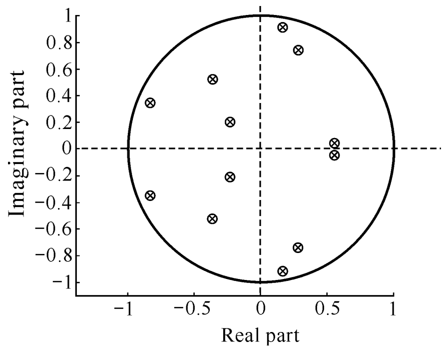

Fig. 3 Pole-zero distributing map of a original individual (○: zeros, ×: pole)

After that, the positions of the zeros and poles (generated in unit circle) are respectively encoded as the elementary operation variables, named as genes; and then the genes are combined to create an individual, i.e., chromosomes. The coding structure of a chromosome is shown below in Fig. 4.

$$b_1\, b_2 \cdots b_M \quad a_1\, a_3 \cdots\, a_N$$

Fig. 4 Chromosome configuration of the parameters' coding

The simulation results also show that the original population created in this way is favorable to the search of the optimal individual to approximate the target DTF. Furthermore, the orders M and N of the pole-zero model are selected as 12. Some things are to be mentioned here. As for the different positions, the complex system response is not in agreement because of the body's different shadow effect. Thus, the orders M and N also need to be different. Here, we don't consider the smooth preprocessing of the original HRTFs; and many details of HRTFs also have no remarkable influence on identifying a sound's location. Thus, the order 12 of M and N are suitable for our study [4, 9].

3) Objective function of individuals' performance Under a logarithmic scale, the magnitudes of the DTFs are fitter for the processing mechanism of sound signal in the human ear; thus the individual performance's assessment and comparisons with the Prony and Yule-Walker are carried out using this kind of scale. In addition, considering the importance of the 900–15 000 Hz range for human hearing's spatial apperception, the magnitudes in this area are much more valued than those outside this range, and the weighted values w_n are 10:1 [3, 5, 6, 9]. The following formula is the objective function of individuals' performance,

$$O_j = \frac{1}{N}\sum_n w_n^2 (\log_{10}|H(f_n)| - \log_{10}|H'(f_n)|)^2 \quad (3)$$

where $H(f_n)$ is the measured DTF, and $H'(f_n)$ is the pole-zero model's magnitude.

4) Individuals' evolutional operation A simple GA has three kinds of evolutional operations: selection, crossover and mutation [8]. The improved GA extends these operations to achieve more effective searching for special problem.

Selection is also entitled reproduction operation. It realizes the individuals' selection through their fitness indicated by the objective function. A simple GA usually uses a basic roulette wheel selection method (RWSM) in the evolutional process. This method uses probability to select individuals based on some measure of their performance. Here, aiming at preventing the premature convergence in the GA, the following linear transformation, which offsets the objective function's fitness, is used prior to fitness assignment,

$$F(r_i) = 2 - S_p + 2(S_p - 1)\frac{r_i - 1}{N_{\text{ind}} - 1} \quad (4)$$

where r_i is the position in the ordered population of individual *i*; $S_p \in [1.1, 2]$ is used to determine the selective pressure and is set to 2 in our application; N_{ind} is the number of individuals and is set to 20 for each sub-group here; *F* is then the individual's fitness after transformation.

Crossover applies the following intermediate recombination designed for real-valued GA,

$$O_1 = P_1 + \alpha(P_1 - P_2)$$
$$O_2 = P_2 + \alpha(P_1 - P_2) \quad (5)$$

where P_1, P_2, O_1, O_2, are respectively the parents before a crossover operation and the filial generations after operation; and α is a scaling factor chosen uniformly at random over the interval [−0.25, 1.25].

Mutation operation use the uniform mutation,

$$O_j = P_j SR\delta \quad (6)$$

where P_j, O_j are respectively the parents before a mutation operation and the filial generations after mutation; $S = \pm 1$ with equivalent mutation probability, set to 0.04 in our program; *R* is half of the variables' bound, determined by the parameter's value generated in a unit circle; $\delta = \sum_{i=0}^{m-1} a_i 2^{-i}$, and $\alpha_i = 1$ with probability 1/*m*, else 0, *m* = 20, and then the precision of the mutation operation can reach $R \times 2^{-19}$.

Figure 5 is a demonstration of the improved GA's convergent trendline. It can be seen from the figure that, the fitness of an optimal individual in the evolutional process takes on a descending trend and ultimately goes flat, which accords with the rule of objective function's minimization. Moreover, the algorithm goes through a full-scale search and basically finds out the optimal solution before the 100th generation; after that the algorithm primarily does local optimization. Therefore, the improved GA has high efficiency in searching, which indicates the rationale for our use of adoptive evolutional strategies and genetic operators.

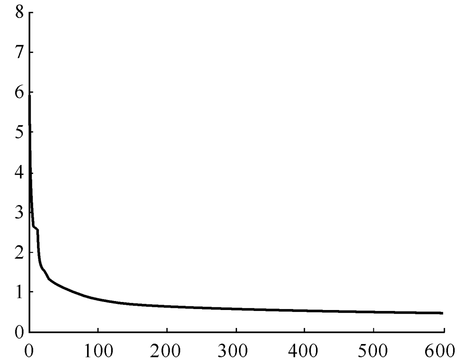

Fig. 5 A demonstration of the improved GA's convergent trendline

Figure 6 shows the improved GA's result at the position (azimuth = 30°, elevation = 0°), where the algorithm of Prony and Yule-Walker apply the functions provided by MATLAB signal processing toolbox.

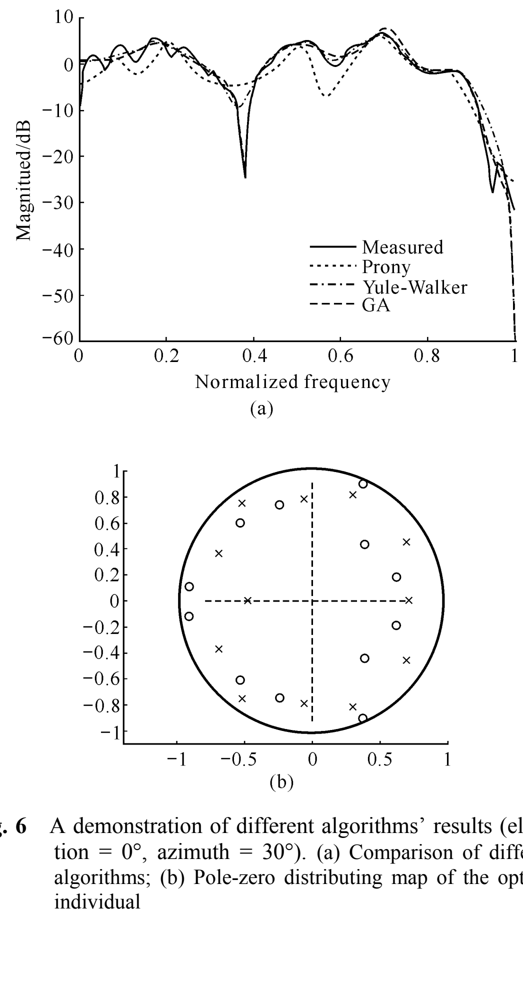

Fig. 6 A demonstration of different algorithms' results (elevation = 0°, azimuth = 30°). (a) Comparison of different algorithms; (b) Pole-zero distributing map of the optimal individual

It can be seen in Fig. 6(a) that, the result of the GA is especially better than that of Prony and Yule-Walker in the 900−15 000 Hz range which is the human hearing's sensitive spatial apperception; meanwhile, based on the log-least square error criterion Eq. (7), GA's error 0.263 8, is better than Prony's 0.458 1 and Yule-Walker's 0.466 9 for the whole spectrum. In addition, considering the GA's randomness, we run the improved GA 20 times and take the average to get the GA's error 0.239 9, also better than that of Prony and Yule-Walker. In Fig. 6(b), the pole-zero distributing map of the optimal individual designed using the improved GA is also given, which shows that the optimal model effectively confines the poles and zeros in the unit circle to guarantee the model's minimum-phase requirement.

Here, we use the following error formulation:

$$\text{error} = \frac{\sum_n (20\log_{10}|H(f_n)| - 20\log_{10}|H'(f_n)|)^2}{\sum_n (20\log_{10}|H(f_n)|)^2} \tag{7}$$

where H is the measured DTF, and H' is the reduction model's DTF.

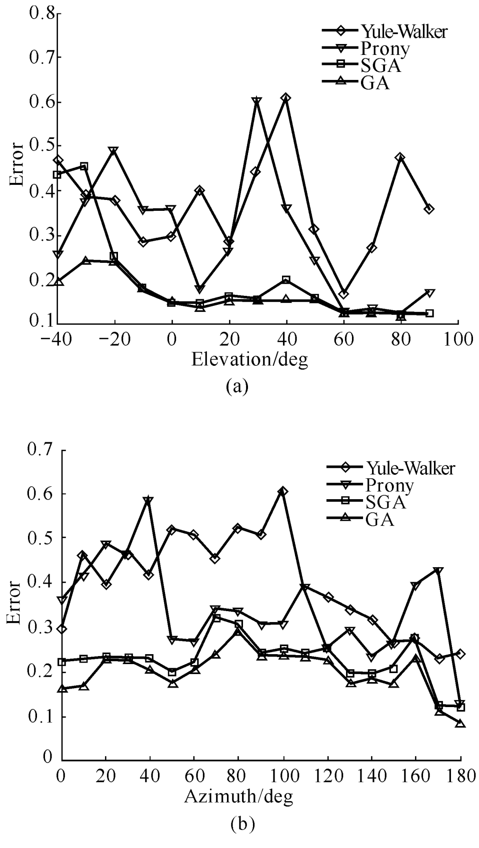

Fig. 7 Results' comparison of different algorithms on vertical and horizontal plane. (a) Vertical plane; (b) Horizontal plane

We also validate some other positions' situations, which is basically in accordance with the above result on the whole. In Fig. 7, the mid-vertical and horizontal plane's results are presented. We can see that, the GA's results noticeably precede Prony and Yule-Walker. The performance of the GA is on the average 39 % better than that of the traditional Prony method, and 46 % better than that of the Yule-Walker algorithm; besides, the two conventional methods' results seriously fluctuate along with the position's movement, while GA displays a steady performance. Further, Fig. 7 also gives the single-group GA's (SGA) result, from which we can see that the strategy of a multi-group insures the algorithm to efficiently search for the optimal solution.

4 Conclusions

In our work, real-valued, multi-group GA is applied to approximate the DTFs' pole-zero model. As seen from the results, this method achieves better effects than the Prony and Yule-Walker based on a weighted log-least square error criterion for the human auditory system. As for future work, we think that, DTFs' similarity and GA's parallelity can be given more importance, which helps to optimize the generation of individuals and parallel the operations for DTFs' approximation. Furthermore, with a view to GA's well optimized performance, better model styles, for example structure model of the outer ear and relevant GA's schemes can be introduced to find a more reasonable model for HRTFs.

Acknowledgements This work was supported by the National Basic Research of China (No. 2002CB312102)

References

1. Blauert J. P., Spatial hearing (revised edition), Cambridge, MA: MIT, 1997
2. Wu Zhen-yang et al., A time domain binaural model based on spatial feature extraction for the head-related transfer function, In J. Acoust. Soc. Am., 1997, 102(4): 2211–2218 (in Chinese)
3. Huopaniemi J., Karjalainen M., HRTF filter design based on auditory criteria, Proc. Nordic Acoustical Meeting (NAM'96), Helsinki, 1996: 323–330
4. Kulkarni A., Colburn H. S., Infinite impulse response models of the head-related transfer function, In J. Acoust. Soc. Am., 2004, 115(4): 1714–1728
5. Durant E. A., Wakefield G. H., Efficient model fitting using a genetic algorithm: pole-zero approximations of HRTFs, IEEE transaction on speech and audio processing, 2002, 10(1): 18–27
6. Blommer M. A., Wakefield G. H., Pole-zero approximations for head-related transfer functions using a logarithmic error criterion, In IEEE Trans. Speech Audio Processing, 1997, 5: 278–287
7. Bill Gardner, Keith Martin, HRTF measurements of a KEMAR dummy-head microphone, Technical Report #280, MIT Media Lab Perceptual Computing, Cambridge, MA: 1994
8. Chipperfield A. J., Fleming P. J. The MATLAB genetic algorithm toolbox, IEE Colloquium on Applied Control Techniques Using MATLAB. Digest No. 1995
9. Kulkarni A., Colburn H. S., Role of spectral detail in sound-source localization, 1998, Nature, 396: 747–749