Automatic Grading System for Mandarin Proficiency Test based on PSO-ANN

Yang Liu
Zhejiang University of Science and Technology
Email: hzliuyang@gmail.com

Chunting Yang
Zhejiang University of Science and Technology
Email: hzyangct@gmail.com

Abstract—**The National Mandarin Proficiency Test(NMPT) to be completed by computer-aided evaluation, not only can remove the human factor to establish uniform standards, but also save time, improve efficiency, accelerate the promotion and popularization of mandarin. In this paper, we focus on the optimization of artificial neural network by PSO for automatic grading the NMPT. A framework of the automatic grading system for mandarin proficiency is presented. The system included three modules which are pronunciation parameter database, evaluation model and automatic grading module. The PSO algorithm is utilized to adjust the connection weights of the selected ANN topology. Experimental results show the method performs better than the exiting methods.**

Index Terms—**automatic grading; mandarin proficiency test; PSO-ANN**

I. INTRODUCTION

Mandarin is the standard language of the modern Chinese, is the communicate tool between various nationalities and areas in our country, so the popularization of mandarin has great significance to the reform and opening and the socialist modernization construction. With the social and economic development, Mandarin, the standard language of the modern Chinese, is great popularized all over China.

Now the mandarin test becomes the national test, and more than three million people take part in National Mandarin Proficiency Test (NMPT) each year in China. And many civil servants, teachers, students and other people start to take part in NMPT. Up to now NMPT is manual. In the process of test, two examiners test one examinee face to face at the same time. The artificial testing has some defects such as so many dialects in our country, the level of examiner is irregular resulting in non-uniform scoring standard, manual control with so much subjective consciousness, and lack of mandarin examiners blocked the rapid development of mandarin. Actually the manual testing has many disadvantages, the following are the main [1][2].

It is inefficiency. When the test is in the way of manual work, many examiners are arranged to the examination rooms. There are two examiners in each room. The examinees that are waiting in the waiting room are call to enter the room in turn. At the same time two examiners only test one examinee and the process may last twenty minutes. Two examiners could test about twenty examinees one day, and the workload of examiners is very heavy. On the other hand, lack of mandarin examiners make the situation worse. These causes a lot of people cannot get the chance of the test.

The testing qualities also face the challenge. There are many dialects in China, and the level of examiner is irregular. That result in non-uniform scoring standard. Otherwise, the test is easy to effect by the subjective consciousness of examiners when the test is processed face to face. Sometime two examiners who are working together may discuss each other. In fact the evaluation should be carried through independently by the examiners.

Attention should be paid to the testing cost. The distribution of costs across the different activities in the test of mandarin is clear. The fee of organization, rent of examination room and wages are the main body. The total cost of testing will cut down at least 50%.

In order to remove the human factors improve efficiency and save the cost, NMPT based on computer-aided is expected urgently. If the mandarin proficiency test is completed by computer-aided evaluation, namely establish standard mandarin pronunciation objective evaluation system, can not only remove the human factor to establish uniform standards, but also save time, improve efficiency, accelerate the promotion and popularization of mandarin. If NMPT based on computer-aided become available, the corpus of NMPT can be obtained. With the data mining technology we can get the attributes of defects mandarin pronunciations. These attributes will greatly contribute to mandarin teaching.

General, there are two types of pronunciation assessment methods, subjective assessment and objective assessment. In subjective assessments the observers' judgment based on rules determines the grade. It reflects the observers' subjective impression on speech quality. Different subjective assessment methods test speech

The research was sponsored by Zhejiang Province Natural Science Foundation of China under Grant No. X106870

doi:10.4304/jcp.5.11.1739-1746

quality on the different focus. There are many methods of subjective assessment, such as mean opinion score (MOS), diagnostic rhyme test (DRT), degradation mean opinion score (DMos), Diagnostic acceptability measure (DAM). Which, MOS is a widely used method of subjective assessment. Its measure the speech quality with the average opinion score of observers. It expresses the speech quality with five levels that are excellent (5 mark), good (4 mark), average (3 mark), poor (2 mark), bad (1 mark). MOS should be a true representation of the speech quality because people are the ultimate recipients of speech.

The advantages of subjective assessment methods are simple and easy to understand. These also are true reflection of the actual speech quality. But in subjective assessment methods there are strict requirements on the assessment condition and process. In order to avoid the perceived bias of individual assessment, subjective assessments should count the statistical results of many observers. So subjective assessment methods are laborious, time-consuming, high cost, poor flexibility, and poor reproducibility, these are also difficult to apply real-time occasions.

To make up for deficiencies in the subjective assessment methods, various objective assessment methods of speech quality have been proposed and applied [3][4]. Objective assessment methods automatically determine speech quality by computer.

According to feature parameters used in speech assessments, objective assessment methods of speech quality has generally experienced the development from the time-domain analysis to the frequency domain analysis, and from the frequency domain analysis to the perception analysis. Based on frequency domain analysis the perceptual domain analysis usually converted speech signal to internal acoustic feature that reflect the psychological characteristics. To some extent the transformation simulate psycho-acoustic characteristics of human and the speech processing in peripheral auditory system and Cochlea of human. The speech feature parameters generated by the processing are considered contains high-level perceptual processing of the nervous system. Now the main objective assessment methods are based on feature parameters of speech perceptual analysis.

The objective assessment methods can be divided into non-reference and reference methods. Reference methods assess the speech quality with the distortion between output speech signal and reference speech signal. On other hand non-reference methods only use output speech signal to assess the speech quality. Objective speech quality metric is a parameter in a kind of feature space. Linear prediction cepstrum coefficients distance (LPCCD), Mel frequency cepstrum distance (Mel-CD), bark spectral distortion (BSD) are typical metrics.

The basic idea of LPCCD is using a linear combination of the past several speech signals sampling to approximate a speech signal sampling. Then a group of predictive parameters are decided by the optimization of difference between speech signal sampling and liner predictive under some criterion. It is appropriate that LPCCD is used as speech distortion metric. But the characteristics of human ear are not considered when LPCCD is used to assess speech signal. Therefore the assessment of LPCCD and subjective assessment may be relatively large deviation.

To different frequencies of sound waves, human auditory sensitivity is different. Actually the relationship between human auditory and the acoustic frequencies is logarithmic. And human auditory perception has the masking effect. According to human auditory perception, the frequency of the acoustic signal is divided by non-uniform. Then some new metrics, such as Mel-CD, BSD, were proposed.[5][6]

Based on cepstral coefficients in the distortion measure and the non-linear frequency characteristics of human auditory perception, Mel-CD is proposed by R. Kublchek. First the frequency axis is transform into scale for Mel cepstral, and then transform into cepstral coefficients obtained in cepstral domain. Mel-CD use weighted sum of squares to define the distortion metrics. Mel-CD may more accurately reflect the human auditory perception on speech, so it has been widely used in the speech recognition and identification.

Bark Spetral Distortion is also one of objective assessments of speech based on human auditory perception. BSD construct transformation model to simulate the human perception mechanism of the speech signal on the basis of human psychoacoustic characteristics.BSD convert speech spectrums into auditory perception spectrum, in the 20 Hz ~ 16 kHz audible region, 24 Barker frequency groups which have different center frequencies are constructed.BSD metric is defined as Barker spectrum Euclidean distance between the original signal and the speech signal. BSD metric simulate the human auditory characteristics with a wide range of hearing.

Mel-CD and BSD are two typical spectrum distortion assessment methods. When they are used to process the speech signal, parts of the human auditory characteristics have been taken into account. Because the frequencies of the speech signal are divided by non-uniform, the assessment results are closer to the subjective assessment.

II. SUBJECTIVE GRADING FOR MANDARIN PRONOUNCING

The smallest unit of Mandarin pronunciation is syllables; all of more than 10,000 Chinese characters are corresponding to only about 1300 syllables. In accordance with the provisions of “Mandarin Proficiency Test level standards", the standard of Mandarin is divided into three levels, each level divided into two grades.

Phonetic error, phonetic defect, intonational deviance and phonetic systemic defects of consonant or vowel are major error in NMPT. Phonetic error means a syllable is misread for another syllable. When a consonant, vowel or tone is misread, the syllable is misread. There are many types of phonetic defect. One of defects is consonant defect. It refer to the oral part of pronunciation is not accurate enough, but not misreading for another

consonant. Vowel defect refers to wrong-shaped mouth or open enough, obviously not listening to a sense of nature; Compound vowels defect are not allowed to place the tongue, not enough movement. Tone defects means the basic tone adjustment and tune trend are correct, but the tone value is obviously low or high, especially the relatively high or low of four tones are obviously inconsistent and so on. When a consonant, vowel or tone is defect, the syllable is defect. Generally certain types of defects are more than 10 times in NMPT, they can determine phonetic systemic defects. That means heavy dialect accent. Intonational deviance are deviance of sentence intonation,or stress.

NMPT consist of four parts that include reading 100 monosyllabic words, reading 100 multiple syllable words, reading a 400-syllable short essay, 3-minute speech.

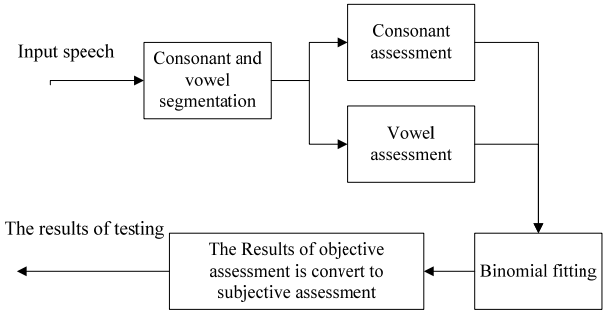
According to the total score, examinee's Mandarin level is assessed.

III. THE PRINCIPLE OF MANDARIN PRONOUNCING GRADING

The smallest unit of Mandarin pronunciation is syllable. It is the basic unit of speech structure.Traditional phonology analyzed a syllable into consonant and vowels. The beginning of a syllable is called as consonant. The consonant letters are used to indicate the sound. There are 21 consonant in Mandarin. The pronunciation length of consonant is relatively stable. This feature is very useful to the objective assessment. A part of the back consonant in a syllable is called vowel. There are 39 vowels in Mandarin.

Mandarin is a tone language, tone is one of the important attributes of Mandarin, and plays an important role in the assessment of Mandarin. In addition to consonant, vowel, syllable also includes tone. Tone is an important component of the syllable. The syllable consists of the same consonant and vowel, if the tone is different the meanings of syllable is not same.

The assessment of Mandarin syllable pronunciation accuracy is contained in two aspects. From the assessment point of view, composition of a syllable phoneme can be assessed with the information of the acoustic channel. On the other hand the tone of a syllable is included in the information of acoustic source. Mandarin tone information is contained in the pitch curve of a syllable, and mainly in the segment of vowel [6].



The process of Mandarin objective grading

Syllable information in these two aspects is independent of each other. They can be processed separately.

The speech assessment principle is shown in Fig.1. Firstly the input speech syllable is segmented into consonant and vowel so as to assess them separately. Because the tone information connote in the vowel, the tone assess with vowel. After consonant and vowel are assessed respectively, binomial fitting method is used to process the results. Then the objective assessment results should be obtained, and finally the results are converted to the testing levels.

IV. A FRAMEWORK OF AUTOMATIC PRONUNCIATION GRADING SYSTEM

A. Particle swarm optimization (PSO)

Particle swarm optimization (PSO) is a population based stochastic optimization technique developed by Dr. Eberhart and Dr. Kennedy in 1995[7], inspired by social behavior of bird flocking or fish schooling.PSO shares many similarities with evolutionary computation techniques such as Genetic Algorithms (GA). Compared to GA, the advantages of PSO are that the PSO requires less parameters and shorter computation time. PSO has been successfully applied in many areas: function optimization, artificial neural network training, fuzzy system control, and other areas where GA can be applied.

PSO is initialized with a group of random particles (solutions) and then searches for optima by updating generations. In every iteration, each particle is updated by following two "best" values. The first one is the best solution (fitness) it has achieved so far. (The fitness value is also stored.) This value is called pbest. Another "best" value that is tracked by the particle swarm optimizer is the best value, obtained so far by any particle in the population. This best value is a global best and called gbest. When a particle takes part of the population as its topological neighbors, the best value is a local best and is called lbest.

In PSO, instead of using genetic operators, each particle (individual) adjusts its "flying" according to its own flying experience and its companions' flying experience. Each particle is treated as a point in a D-dimensional space. The *i*th particle is represented as $X_I = (x_{i1}, x_{i2}, \ldots, x_{iD})$. The best previous position (the position giving the best fitness value) of the *i*th particle is recorded and represented as $P_I = (p_{i1}, p_{i2}, \ldots, p_{iD})$. The index of the best particle among all the particles in the population is represented by the symbol g. The rate of the position change (velocity) for particle *i* is represented as $V_I = (v_{i1}, v_{i2}, \ldots, v_{iD})$. The particles are manipulated according to the following equation:

$$v_{id} = w * v_{id} + c_1 * \text{rand}(\ ) * (p_{id} - x_{id}) + c_2 * \text{Rand}(\ ) * (p_{gd} - x_{id}) \quad \text{(la)}$$

$$x_{id} = x_{id} + v_{id} \quad \text{(lb)}$$

where c_1 and c_2 are two positive constants, rand( )and Rand( )are two random functions in the range [0,1], and w is the inertia weight. Equation (la) is

used to calculate the particle's new velocity according to its previous velocity and the distances of its current position from its own best experience (position) and the group's best experience. Then the particle flies toward a new position according to equation (lb).[8]

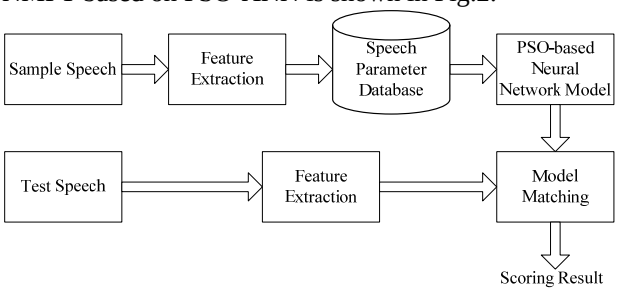
The performance of each particle is measured according to a predefined fitness function, which is related to the problem to be solved. The inertia weight w is employed to control the impact of the previous history of velocities on the current velocity, thereby influencing the trade-off between global (wide-ranging) and local (nearby) exploration abilities of the "flying points." A larger inertia weight w facilitates global exploration (searching new areas) while a smaller inertia weight tends to facilitate local exploration to fine-tune the current search area. Suitable selection of the inertia weight w can provide a balance between global and local exploration abilities and thus require fewer iterations on average to find the optirnum. In this paper, an analysis of the impact of this inertia weight together with the maximum velocity allowed on the performance of PSO is given, followed by experiments that illustrate the analysis and provide some insights into optimal selection of the inertia weight and maximum velocity allowed.

B. Artificial Neural Network (ANN)

An artificial neural network (ANN), usually called "neural network" (NN), is a mathematical model or computational model that tries to simulate the structure and/or functional aspects of biological neural networks. It consists of an interconnected group of artificial neurons and processes information using a connectionist approach to computation. In most cases an ANN is an adaptive system that changes its structure based on external or internal information that flows through the network during the learning phase. Neural networks are non-linear statistical data modeling tools. They can be used to model complex relationships between inputs and outputs or to find patterns in data.

C. The Framework of the System

The framework of automatic grading system for NMPT based on PSO-ANN is shown in Fig.2.



A framework of automatic pronunciation grading system

1) Establish pronunciation parameter database. Mandarin proficiency test system grades the examinees' pronunciation based on comparing similarity between the pronunciation characteristic of the examinee and the standard pronunciation in the pronunciation parameter database. Therefore the establishment of the pronunciation parameter database is a most foundation link in the entire system. In order to achieve ideal effect, generally selects the person whose pronunciation is extremely standard or the pronunciation material to train the pronunciation parameter database.

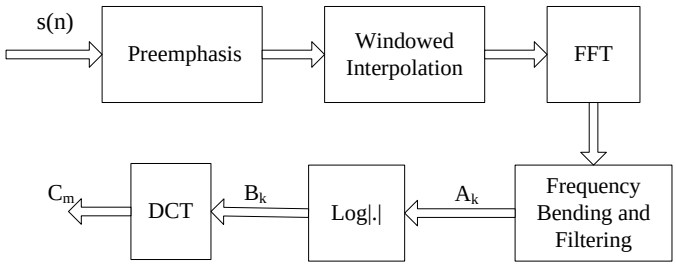
2) Establish evaluation model based on artificial neural network. Based on the Mel parameter extraction, calculate distorted distance between sample speech and test speech, and combine with the subjective evaluation criteria establish the objective prediction model based on artificial neural network.

3) Automatic scoring. Collect examinee's pronunciation data, preprocess and extract feature. Based on the well trained unit module in the pronunciation parameter database, divide speech to small units which used for calculating the pronunciation quality by compulsory alignment. According to the characteristic of the mandarin pronunciation, the small unit is generally a phoneme. Compare the examinee's phonation with the pronunciation in the parameter database, according to the speech recognition principle, use evaluation model based on artificial neural network quantification index, finally output the result referring to the scoring standard. This is the most important part in the system.

V. OBJECTIVE EVALUATION MODEL BASED ON PSO-ANN

A. Feature Parameter Extraction

The feature parameter extraction is the most important part in the pronunciation quality objective evaluation. The human's perception to the speech signal contains lots of psychoacoustics knowledge. Based on psychoacoustic and uses objective evaluation distortion measure which reflect human's sensation characteristic, can obtain better correlation between subjective evaluation result and objective evaluation result, only then it be possible to use these parameters to evaluate the pronunciation accuracy objectively. Mel is a unit to measure pitch in the psychoacoustic and to describe the subjective sensation of human ear to sound frequency. Mel Cepstrum distortion measure is a bending frequency spectrum. It fully reflects the human ear's non-linear perceptual characteristics to the frequency and the frequency analysis and spectrum synthesis feature of the human ear when it hears complex sound. Mel Cepstrum is the most commonly used feature parameters in the speech signal processing. The flow of Mel Cepstrum extraction shows in Fig.3. [10,11]



The flow of Mel Cepstrum extraction

Mel Cepstrum coefficient speech analysis is based on the cepstrum analysis. After image deconvolution, the linear frequency scale needs to be mapped to the Mel frequency scale, and then through a group of triangle bandpass filter, bandwidth of each bandpass filter is uniform in the Mel frequency scale. Let Mel is the Mel frequency, the unit is Mel, f is the linear frequency, the unit is Hz. The mapping relation between the Mel frequency and the linear frequency is approximately:

$$\mathrm{Mel} = \begin{cases} \mathrm{f}, & \mathrm{f} \le 1000\mathrm{Hz} \\ 1000\log_2\left(1+\frac{\mathrm{f}}{1000}\right), & \mathrm{f} > 1000Hz \end{cases}$$

After frequency bend processing, let power spectrum $\overline{\mathrm{X}}(\mathrm{f})$ through Mel measure triangle bandpass filter, obtain energy weighted sums $\mathrm{A_k}$ when it pass through every digital filter, and solve its logarithm value$\mathrm{B_k}$. For discrete cosine transform obtain MFCC coefficient $\mathrm{C_1} \sim \mathrm{C_M}$:

$$\mathrm{C_m} = \sum_{\mathrm{k=1}}^{\mathrm{M}} \mathrm{B_k}\cos\left[\mathrm{m}\left(\mathrm{j}-\frac{1}{2}\right)\frac{\pi}{\mathrm{M}}\right], \mathrm{m} = 1,2,\ldots,\mathrm{M}, \quad \mathrm{M} = 12$$

By the MFCC of the test pronunciation and sample pronunciation, the distortion of each test pronunciation is calculated, and obtains the square sum of distortion measure is:

$$\mathrm{d_{Mel}(n)} = \sum_{\mathrm{m=1}}^{\mathrm{M}} (\mathrm{C'_{n,m}} - \mathrm{C_{n,m}})^2, \mathrm{n} = 1,2,\ldots,\mathrm{N}$$

n is the frame number, it decided by the length of the pronunciation document. $\mathrm{C'_{n,m}}$ and $\mathrm{C_{n,m}}$ represent nth frame m-dimensional MFCC coefficient of test speech signal and sample speech signal respectively. Thus, obtains Mel cepstrum coefficient distorted distance of each frame. Finally, calculate the arithmetic mean of Mel cepstrum coefficient distortion distance of each frame of pronunciation document, obtains Mel cepstrum coefficient distortion distance of this distorted document, this is the Mel cepstrum coefficient distortion of the test pronunciation.

When uses neural network system model for objective test speech quality evaluation, select two feature parameters as the neural network input. One kind is the Mel cepstrum feature parameter, namely

$$\mathrm{in} = \sum_{\mathrm{m=1}}^{\mathrm{M}} (\mathrm{C'_{1,m}} - \mathrm{C_{1,m}})^2, (\mathrm{C'_{2,m}} - \mathrm{C_{2,m}})^2, \ldots, (\mathrm{C'_{n,m}} - \mathrm{C_{n,m}})^2$$

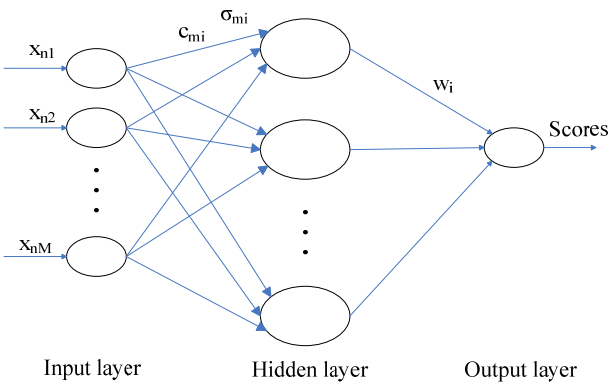
as neural network n input. The other kind is the Mel cepstrum difference feature parameter, namely

$$\mathrm{in} = \begin{bmatrix} \mathrm{C'_{1,1}} - \mathrm{C_{1,1}} & \cdots & \mathrm{C'_{1,M}} - \mathrm{C_{1,M}} \\ \vdots & \ddots & \vdots \\ \mathrm{C'_{N,1}} - \mathrm{C_{N,1}} & \cdots & \mathrm{C'_{N,M}} - \mathrm{C_{N,M}} \end{bmatrix}$$

as neural network n m-dimensional input.

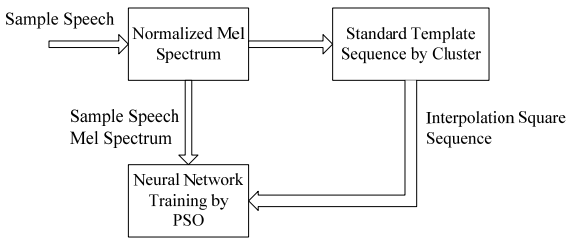
B. *Objective Evaluation Model based on Artificial Neural Network*

The speech quality objective evaluation, in essence, can be interpreted as a super-surface fitting in high-dimensional space using neural network. The input vector of the automatic pronunciation system is$[\mathrm{x_{n1}}, \mathrm{x_{n2}}, \ldots, \mathrm{x_{nM}}]$, it is the output speech signal's m-dimensional Mel cepstrum feature parameter of the nth frame. Through the neural network, obtains the normalized output. Then based on the negative correlation between the outlet of the neural network and the subjective score, transform it to objective evaluation estimate value between the [0,100] through the linear relations.

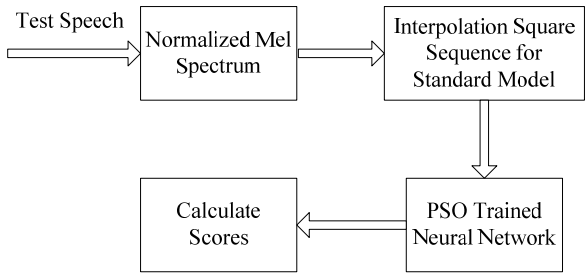


Mandarin proficiency objectively evaluating based on the neural network

Establish the neural network to objectively evaluate the mandarin proficiency include two stages: learning stage (as shown in Fig.5) and network scoring stage (as shown in Fig.6). In the learning stage, the sufficient pronunciation sample is selected to training neural network, and the learning result saves as the weight form in the neural network structure. In the network scoring stage, when test pronunciation signal pass through the neural network, the neural network which is well trained and having certain generalization performance match feature by interpolation and extrapolation adaptively, and according to the best match result and the accuracy subjective evaluation criteria, obtain subjective evaluation result of the test speech predicting by the objective evaluation.



Learning stage



Network scoring stage

MFCC spectrum feature parameter corresponding to spectrum distortion measure usually measured by Euclidean Distance, namely:

$$D(X'_{k,n}, X_n) = \left[\sum_{m=1}^{M} (x'_{k,nm} - x_{nm})^2\right]^{\frac{1}{2}}, \quad n = 1,2,\dots,N, k = 1,2,..,K$$

In order to improve the anastomosis between Euclidean Distance and the subjective sensation judgment, interpolation square of each feature component multiplied a weighted coefficient, obtains:

$$D_w(X'_n, X_n) = \left\{\sum_{m=1}^{M} [w_m \cdot (x'_{nm} - x_{nm})^2]\right\}^{\frac{1}{2}} \quad n = 1,2,\dots,N$$

This kind of weighted Euclidean Distance reflects the difference between each dimension of the feature vector, is closer to the real hearing characteristic, to a certain extent can make up MFCC itself not to be able to describe the relative important of each dimension. In m-dimension space, the iso-surface of $D(x'_n - x_n)$ form a positive hypersphere in which the center is zero-vector and the radius is $D(x'_n - x_n)$.The iso-surface of $D_w(x'_n - x_n)$ when its weight is nonnegative form a positive super ellipsoid in which the center is zero-vector and the radius is $D_w(x'_n - x_n)$. Furthermore, if the concept of single-layer perceptron was extended to the general single-output nodes feedforward neural network, equivalent to turn weighted distance function $D_w(x'_n - x_n)$ to more complex nonlinear function which manifests by the neural network. Their iso-surface of neural network-based distance function in m-dimensional space will be surrounded by some of the closed surface which complexity of the zero-vector can be adjusted randomly. After adequate self-adaptive training, this closed surface can effectively differentiate various distortion conditions, and can realize the best adaptive approximation to the pronunciation distortion judgment characteristics of the auditory system.

C. PSO algorithm training Artificial Neural Network

Optimization computation methodologies have been applied to three main attributes of neural networks: network connection weights, network architecture (network topology, transfer function), and network learning algorithms. PSO is a promising method to train ANN. It is faster and gets better results in most cases.

The particle will be a group of weights. After encoding the particles, we need to determine the fitness function. For the classification problem, we feed all the patterns to the network whose weights is determined by the particle, get the outputs and compare it the standard outputs. Then we record the number of misclassified patterns as the fitness value of that particle. Now we can apply PSO to train the ANN to get lower number of misclassified patterns as possible. There are not many parameters in PSO need to be adjusted. We only need to adjust the number of hidden layers and the range of the weights to get better results in different trials.

PSO algorithm training Artificial Neural Network can be summarized as follows:

Step1. Define the network structure (shown as Fig.4), parameters of PSO and the fitness function.

In formula (1a), the number of particles is 30, search space limit Vmax belongs [-2.0, 2.0], acceleration factors c_1=c_2=2.

Step2. Each PSO starts with a group of networks whose connection weights are randomly initialized. The ANNs are tested and each receives a score. The PSO keeps track of the global best configuration and each ANN's individual best configuration. A configuration for an ANN with n connections can be considered as a point in an n+1 dimensional space, where the extra dimension is for the reinforcement score. After every round of testing, the PSO updates the connection weights of the ANNs' according to the following equations:

$$\mathbf{v}_{t+1} = c_{imr}\mathbf{r}_1 * \mathbf{v}_t + c_{cgn}\mathbf{r}_2 * (\mathbf{x}_{pb} - \mathbf{x}_t) + c_{scl}\mathbf{r}_3 * (\mathbf{x}_{gb} - \mathbf{x}_t)$$

$$\mathbf{x}_{t+1} = \mathbf{x}_t + \mathbf{v}_{t+1}$$

where position and velocity vectors are denoted by $\mathbf{x}$ and $\mathbf{v}$, respectively. The $\mathbf{r}_i$ represents vectors where each element is a new sample from the unit-interval uniform random variable. Personal best, $\mathbf{x}_{pb}$, is the point in the solution-space where that particular ANN received its highest score so far. Global best, $\mathbf{x}_{gb}$, is the point with the highest score achieved by any ANN of this PSO. The three constants, c_{imr}, $c_{cgn}$and c_{scl}, allow the adjustment of the relative weighting for the inertial, cognitive, and social components of the velocity, respectively. After exhausting its allotted training iterations, the PSO reports the global best to the administrator. If/when it is allocated additional training iterations, the PSO resumes training from exactly where it left off.[12]

Step3. Identify the personal best fitness value and update the corresponding position for each particle; identify the global best fitness value and update its position.

Step4. Update the velocity and the position for the whole particle swarm according to equation (1a) and (1b).

Step5. If the stopping condition is not satisfied, go to step 3. Otherwise, terminate the iteration and obtain the best weight setting from the global best solution.

D. Experimental Results

The performance of the objective assessment system was measured by the correlation between the objective evaluation results and the ideal value.

Generally the correlativity of the subjective and objective assessment results can be represented by the Pearson coefficient. The correlative coefficient ρ and deviation δ can be calculated using the following equations.

$$\rho = \frac{\sum_{i=1}^{M}(S_O(i)-\overline{S_O})(S_S(i)-\overline{S_S})}{\sqrt{\sum_{i=1}^{M}(S_O(i)-\overline{S_O})^2 \sum_{i=1}^{M}(S_S(i)-\overline{S_S})^2}}$$

$$\delta = \sqrt{\frac{\sum_{i=1}^{M}(S_O(i)-S_S(i))^2}{M-1}} \quad .$$

Where M is the number of distortion speech, and $S_O(i)$ and $S_{OS}(i)$ are the results of subjective assessment and objective assessment of the i^{th} distortion speech respectively. $\overline{S_O(i)}$ and $\overline{S_S(i)}$ are the average results of subjective assessment and objective assessment of the i^{th} distortion speech respectively.

In this paper all of the training samples and test samples are from the NMPT database of Zhejiang Province Mandarin Professional Testing Center.

TEST DATA OUTPUT COMPARISON

Sample	The ideal output	ANN method	PSO-based ANN method
1	0.8234	0.7568	0.8356
2	0.6753	0.5534	0.6654
3	0.8792	0.8612	0.8802
4	0.7658	0.8513	0.7703
5	0.8250	0.7765	0.8241
6	0.8912	0.8200	0.8897
7	0.6938	0.7356	0.7032
8	0.7324	0.7821	0.7589
9	0.7600	0.7721	0.7512
10	0.8867	0.8789	0.8743

We used traditional ANN methods and improved ANN methods which based on PSO to training the same group data, the results of the algorithms performance comparison were as follows:

(1) To achieve the same error, the iteration times of the traditional ANN method is 4987, and the iteration times of the PSO-based ANN is 3521.Obivouslly, PSO-based method can reduce the iteration times significantly.

(2)Tested by ten samples, the test data output were also different by the two algorithms.

As can be seen from Table 1, the assessment results obtained by PSO-based ANN method were closer to the ideal output value and more precise.

VI. CONCLUSION

In this paper, a PSO-based ANN algorithm is proposed to automatically grading the learning in Internet. Basically, the PSO algorithm is utilized to adjust the connection weights of the selected ANN topology. Taken mandarin learning as example, we introduced the PSO-based ANN algorithm to grading mandarin learning, the experimental results shown it's an effective method.

The above theory has been applied to the information management system of mandarin test (IMSMT).Zhejiang province mandarin test center is the first user of the IMSMT. The Hangzhou city mandarin test center, Yiwu city mandarin test center and Hangzhou normal university are the new users. They all give a high appraisement .About 7140 examinees finish their mandarin test with the IMSMT. The system has passed Chinese Ministry of Education appraisal in 2009.

ACKNOWLEDGMENT

Appreciation Zhejiang province mandarin test center and Hangzhou mandarin test center support for the project.

REFERENCES

[1] Chunting Yang, Yang.liu, and Zhigang Cheng, "A Framework for the Information Management System of Mandarin Test", The Proceeding of IEEE IUCE 2009. Chengdu, pp. 164–167, May 2009.

[2] Yang Liu, Chunting Yang, and Weifeng Ma, "Automatic Pronunciation Scoring for Mandarin Proficiency Test Based on Speech Recognition", The Proceeding of IEEE IUCE 2009. Chengdu, pp. 168–172, May 2009.

[3] H Franco, L Neumeyer, Y Kim, O Ronen ,"Automatic pronunciation scoring for language instruction," IEEE International Conference on Acoustics, Speech, and Signal Processing, 1997 vol.2,pp.1471-1474,1997

[4] L Neumeyer, H Franco, V Digalakis, M Weintraub, "Automatic scoring of pronunciation quality,"Speech Communication, Volume 30, Issues 2-3, pp. 83-93, February 2000

[5] Lee Yu- min, Lee Lin- shan, "Continuous hidden Markov models integrating transitional and instantaneous features for mandarin syllable recognition", Computer Speech and Language, Vol.7, pp. 247- 263,1993.

[6] Ying Y, Xu S, "A fast method of pitch detection for Chinese four tones recognition", Proceeding of ISCP' 93 , Beijing, Oct 1993.

[7] Kennedy, J. and Eberhart, R. C. Particle swarm optimization. Proc. IEEE int'l conf. on neural networks Vol. IV, pp. 1942-1948. IEEE service center, Piscataway, NJ, 1995.

[8] Shi, Y. and Eberhart, R. C. Parameter selection in particle swarm optimization. Evolutionary Programming VII: Proc. EP 98 pp. 591-600. Springer-Verlag, New York, 1998.

[9] H Franco, L Neumeyer, Y Kim, O Ronen ,"Automatic pronunciation scoring for language instruction," IEEE International Conference on Acoustics, Speech, and Signal Processing, 1997 vol.2,1997,pp.1471-1474

[10] Stevens S, Volkman J. The relation of pitch of frequency: a revised scale [J]. Am J Psychol, 194, 53:329-353.

[11] Zwicker E, F lottorp G, Stevens S. Critical bandwidth in loudness summation [J]. J Acoust Soc Am, 1947, 19: 90-119.

[12] Matthew Conforth and Yan Meng,An Artificial Neural Network Based Learning Method for Mobile Robot

Localization, Robotics, Automation and Control,2008,pp,103-112
[13] L Neumeyer, H Franco, V Digalakis, M Weintraub, “Automatic scoring of pronunciation quality,”Speech Communication, Volume 30, Issues 2-3, February 2000, pp. 83-93
[14] Horacio Franco, Leonardo Neumeyer, Vassilios Digalakis and Orith Ronen,“Combination of machine scores for automatic grading of pronunciation quality ,“Speech Communication, Volume 30, Issues 2-3, February 2000, pp. 121-130
[15] Franco H, Abrash V,Precoda K,Bratt H,Rao R, Butzberger J,Rossier R , Cesari F,”The SRI EduSpeakTM system: Recognition and pronunciation scoring for language learning,”Proceedings of InSTILL 2000. Dundee: University of Abertay
[16] Christa van der Walt1,Febe de Wet,Thomas Niesler,”Oral proficiency assessment: the use of automatic speech recognition systems,” Southern African Linguistics and Applied Language Studies 2008, 26(1): 135 – 146

Yang Liu was born in Wuhan, P. R. China, on June 4, 1978. Liu received her BS in 2000 and MS in 2003 from Huazhong Normal University.
She is a full-time lecturer of computer science and engineering at the Zhejiang University of Science and Technology. Her research interests are in swarm intelligent, information processing.

Chunting Yang was born in Qiqihar City, P. R. China, on January 16, 1964. Yang received a BS in 1986 from Nanjing University of Aeronautics and Astronautics. Then he received his MS in 1991 and PhD in 1996 from the Southeast University and Zhejiang University. He is an associate professor of computer science and engineering at the Zhejiang University of Science and Technology. His research interests are in computer image processing, computer vision and speech signal processing.
Dr. Yang is the member of a council of Zhejiang Computer Society and the member of a council of Zhejiang Electronics Society