# PRINCIPLE OF INFORMATION DIFFUSING RELEVANT TO FUZZY SPEECH SOUND RECOGNITION

HUANG CHONGFU* WANG JIADING**

*Department of Mathematics,
Beijing Normal University,
Beijing, China
**Geological Hazards Institute,
Gansu Academy of Science,
Lanzhou, China

ABSTRACT. In this paper, based on the existence of transition information of fuzzy phenomenon, the principle of information diffusing is advanced. In the light of it, we can reasonably diffuse fuzzy information to grasp the regulation of the phenomenon. With the method of injecting information from central point and diffusing it layer by layer, we are able to recognize the fuzzy speech sound that is different person's speech sound.

## 1. INTRODUCTION.

The great majority of objects all over the world are not absolute, isolated, unchangeable and eitheror. On the contrary, they are changeable, connective each other, movable and both. When we talk about something displaying distinctness and stationariness, we are speaking in relative terms. Generally, almost everything is vague and motional. For understanding the world, we use the ideal model being easy and crystal-clear to explain the ambiguous object gradually. And sometime, we use the method which is comparatively static to describe the moving body.

In order to recognize the world, we have to analyse the information collected from nature. By modern technology, the information can be got in digital form. Therefore, it is better to oversee datum in point of view of information. that is,when you handle datum, you ought to remind yourself that the datum are information if they come from nature.In many cases,the view of information can help us to grasp the law of nature.

The primitive emitter of information is objective world. The information after appears in digital form. So, it would be efficacious to analyse datum with the help of the characters of the emitter.

Before the concept of fuzzy set was set up, scientists had suspected the absoluteness of a data which has a actual numerical value. In the random theory, one data can describe nothing. We must put it in the random physical model for

*Key words and phrases.* Information diffusing, numeral information, information distribution, speech sound, recognition.

getting the random nature,and consider the relationship with other datum. Only this way,one data can be used.

Similarly, in the theory of fuzzy set, we should not believe one data is absolute. We ought to let it go into the fuzzy physical model for gaining transitive property, and think the connection with other datum.Like this,one data is very useful.

In the current methods of fuzzy mathematics, a little of fuzzy transition information is utilized. These methods usually touch on two sides only,which is this side and that side. They are short of expressing the total of transitive information. This paper try to do something for it, So the principle of information diffusing is proposed. From the influence of many dots point of view, this principle is a bit similar to set value statistics and gratation statistics.

Human speech sound is a kind of complex fuzzy phenomena. Only when the transition of information is considered in higher level, there is the possible to build the model used to recognize general speech sound. The effective model would help us to dispel the impact of sex and frequency when the samples are matched. It is the same as that one can understand the meaning of the speech sound of other , and sex or frequency can not influence the meaning. The objective was attained.

## 2. SIMPLE DEPICTION OF TRANSITION OF NUMERAL INFORMATION.

The main property of the fuzzy phenomenon is that there is stronger transition for the information concerned. Earthquake phenomenon is a typical example which is not only a random phennomenon but also a fuzzy phenomenon. The numeral information obtained from the phenomenon possesses both randomness and fuzziness. The randomness lies in that we have not totally known every factor which affects earthquake occurrence. The same condition which can be controlled often brews different earthquake in time, place and intensity. Sometimes, the difference is very great. The fuzziness lies in that we have to depend on analyzing historical earthquake material to sum up the experiences and lessons in many fields, because we have not totally grasped the law of earthquake. This kind of material is always a limited number and incomplete. We often lack the historical earthquake memorandum, especially in a certain district. For some types of earthquake, we have not records to directly explain which will happen if the earthquake occurs, but we have to know it because this earthquake affection is important. In this case, using the fuzziness of records, we would know something.

The fuzziness of earthquake record is the transition of the datum of earthquake. The simple describing method firstly arose in the estimation of earthquake epicentral intensity with magnitude given.

There are 134 records in a district,and every record contains two factors which are magnitude M and epicentral intensity I. If using the classical statistical method to find the relationship of M and I, we could not get ideal solution since the records of weak earthquake is greatly more than those of strong earthquake.

After analyzing, we discover that this topic is not only involving random, but do fuzzy. There are a few of records in the important range of earthquake. They are so little, incompleteness is very serious. However, the relationship between $M_1$ and $I_1$ must implicate the commection of $M_2$ and $I_2$, and so on. That is saying, the information can be transferred.

For describing this, the simplest method is to transfer the information given to an attractor according to the distance.

For example , we know magnitude M= 5.75,epicentral intensity I= $VIII$, we want to know when $M_1 = 5.5$ and $M_2 = 5.8$, what possibility is will for $I_1 = VIII.$ and $I_2 = VIII.$

It has been shown in practice that it is perfectly possible to do it by using linear transferring formula (1).

$$\text{Poss}(VIII/M_i) = Q_i = 1 - \frac{|M - M_i|}{\Delta_i} \tag{1}$$

where $\Delta_i$ depends on how many datum will be used, and it is a constant. For estimating epicentral intensity I by these 134 records, we get $\Delta = \Delta_i = 0.3$, so we obtain

$$\left.\begin{aligned} Q_1 &= 1 - \frac{|5.75 - 5.5|}{0.3} = 1 - 0.83 = 0.17 \\ Q_2 &= 1 - \frac{|5.75 - 5.8|}{0.3} = 1 - 0.17 = 0.83 \end{aligned}\right\} \tag{2}$$

Now, we know when magnitude is $M_1$ ,the possibility of $I_1 = VIII$ is smaller, it is Poss(I= $VIII/M_1 = 5.5$) = 0.17, and when $M_2$, is greater, Poss(I= $VIII/M_2 = 5.8$) = 0.83.

Certainly, because random exits,we must totally process all 134 records by adding up possibility and standardizing the result to get a reliable relationship between M and I.

Equation (1) is defined as information distribution formula, $M_i(i = 1, 2, \cdots, n)$ are called control points. The all information of these 134 records are totally distributed to these points. The result of equation (2) shows that every record is considered that the quantity of a record information is 1 which is absorbed by the two points which is close to the record. This kind of distribution method is called two points distribution technique. $\Delta$ is the distance between two points.

If there are many datum, and they are very small, the precision will be higher. Of course, epicentral intensity I can also be recognized by using general statistical technique, but when it comes to non-uniformity of data, the recognition accuracy will be not very good. On the other hand, it is not simply described for the changing case of intensity among varied magnitude by classical statistical method. Therefore, we are still using information transition technique, because the fuzzy nature of data is not holly disappear whenever.

The digit information transition property is analyzed by Equations (1) and (2), it is practical but too simple to satisfy some complex problems. In order to perfecting this method, some scholars try to find better way of information distribution between two points, where mainly, there are a second root form and a second algebraic form. It is apparent that the 2nd root form don't satisfy the rules of information transition. And other forms have a certain waviness. It is difficult to choose optimum form.

## 3. THE PRELIMINARY STUDY ON THE WHOLE UNIVERSE INFLUENCE OF TRANSITION PROPERTY OF INFORMATION.

Within the model of information distribution in two points , the effects to other points may be handled by the statistical and normalizing processes, but it is not direct and clear. Then, scholars tried to expanse the sphere of two points and process information in a expanding universe.

One of the most simple thought is that supposing a formula relatived with distance. Normal distribution is a example:

$$Q(x) = \exp\left(-\left(\frac{x-a}{b}\right)^2\right) \tag{3}$$

where, x-the distance between effected points and fallen points, a, b— constant. Our analysis shows that if one wants a,b to have obvious physical meaning and to satisfy the precision of pattern recognition, there isn't a simple way to determine them.

In the evolution of sandy soil liquefaction potential[2], author, based on the physical phenomenon of molecule diffusion, had introduced a information diffusion formula

$$Q = \begin{cases} \min\{1, (1/2\pi)\ln(2a/r)\} & r \in (0,1] \\ 0 & r \notin (0,1] \\ 1 & r = 0 \end{cases} \tag{4}$$

where, r-the distance function, a-radius of studied universe. But it is difficult to determine the radius of studied universe a. There is another formula was used in the risk analysis of earthquake in Beijing, Tianjin and Tangshan.

It is very difficult to find the best model,but we have found better one. It is the information diffusing technique as follows.

## 4. THE PRINCIPLE AND TECHNIQUE OF INFORMATION DIFFUSING.

Relativity is able to denote the relationship among different points of a random record in random theory. While in the fuzzy information theory, the influence among different points is determined with their information interference, which is information from point A effecting point B, so do converse case.

The principle of information diffusing is that: there are several points $x_1, x_2, \cdots, x_n$ in range between point A and point B.The information which is obtained in point A can be pass to point $x_1$, with the attenuation. again, its parts must pass to point $x_2$, so and so, and to point B.

In fact, if point $x_1$ is very near point A , the information given to point A is almost given to point $x_1$ in the view of fuzzy sets. Therefore, the parts of the information from A must pass to point $x_1, \cdots$, and to point B.

In accordance with the above information diffusing , some special diffusing method is established. In this paper, we shall introduce the technique of falling information from central point and diffusing layer by layer.

Suppose $x_1, x_2$, is two points for absorbing information at two side of point x where information had fallen. Let x is the central point of $x_1$ and $x_2$. If we fall information $Q = 1$ on point x , then $x_1, x_2$ will get information $Q_1 = 0.5$ and $Q_2 = 0.5$ using the method of information distribution as formula (1) and (2).

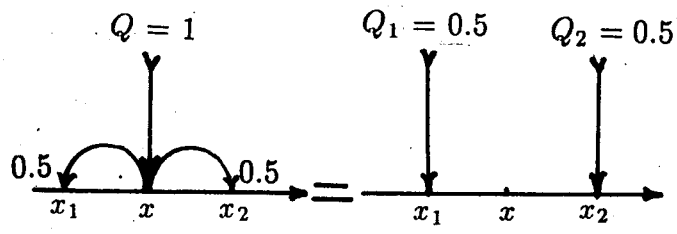



Fig 1 The shift both information distribution and falling

As shown Fig.1,falling $Q = 1$ on x is approximately equal to falling $Q = 0.5$ on $x_1, x_2$ respectively.

Let's choose a point $x_0$ in left side of x and take a point $x_3$ in right side of $x_2$. Now $x_0, x_3$ will obtain information $Q = 0.25$ from $x_1, x_2$ respectively. While x obtain $Q = 2 * 0.25 = 0.5$ from $x_1$ and $x_2$. That is approximately equal to information $Q = 0.25, 0.5, 0.25$ falling from $x_0, x, x_3$ respectively as shown in Fig.2.

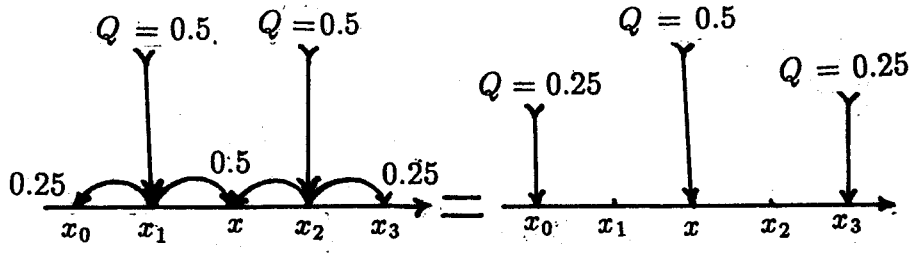



Fig 2 The shift both information distribution and falling

With the increase of absorb points in two vector directions, information can be diffused out. This method is call cential point injecting and layer by layer diffusing, because it is similar to a mechanics model of undergoing force in center point and diffusing layer by layer.

If we put a force (its magnitude $P = 1$) in a point x on a long plate $L_1$, the bearing points $x_1, x_2$ under the long plate $L_1$ will undergo a supporting force (its magnitude $P = 0.5$) respectively. If the two bearing points are sustained by another long plate $L_2$ with bearing points $x_0, x, x_3$ under points $x_1$ and $x_2$ as shown in Fig. 3, and $x_1, x_2$ are the central points of them , $x_0, x, x_3$ will get force $P = 0.25, 0.5, 0.25$ respectively. Of course, with the increasing of number n of bearing points, the force must be diffused . No matter what value n is, the result of force can be calculated by using recurrence formula. when $P = 1$ on x of $L_1$, the last result of diffusing value is called diffusing coefficient which is denoted by vector $\vec{C}$. If $n = 11$, diffusing coefficient is :

$$(0.003, 0.016, 0.054, 0.121, 0.193, 0.226, 0.193, 0.121.0.054.0.016.0.003) \quad (5)$$

Because of the symmetryy above vector $\vec{C}$, it can be written by its half as following

$$C_6 = (c_1, c_2, \cdots, c_6) = (0.226, 0.193, 0.121, 0.054, 0.016, 0.003) \quad (6)$$

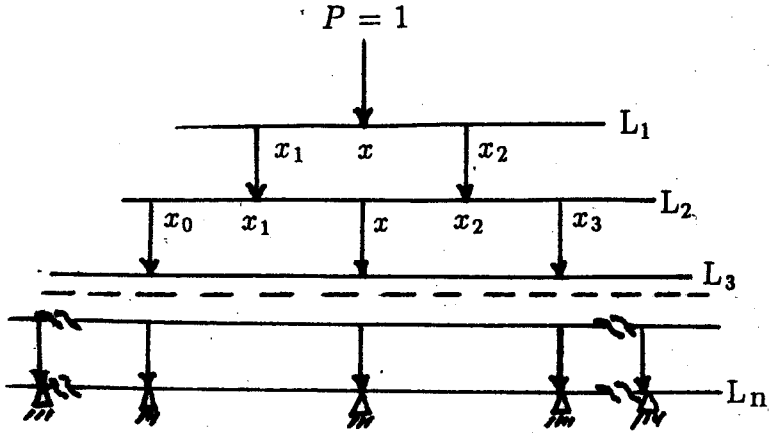



Fig 3 Information model of fallen force from central point

When $n = \infty$, we will get $C_\infty = 0$. It satisfies the principle of intensity of pressure. In other words, if P is certain value and bearing area is infinite, intensity of pressure is 0. But the bearing area in many cases is finite, and intensity of pressure is not 0. In analysis of data information, we are interested in influence of information between different points, and the universe is finite, hence diffusing coefficient can't be 0.

## 5. INFORMATION DIFFUSING PRINCIPLE AND PRONUNCIATION RECOGNITION.

As today, it has reached a top of tree as a special pronunciation recognizer. Using RTSRS system produced by Acoustics Institute, Academia Sinica, the correct recognition ratio has reached 99% . It has been successful to produced a typewriter putting pronunciation. All of these works have been finished by using the technique of spectrum analysis. Of course, most of scholars do their utmost to improve recognition accuracy. However, the technique of spectrum analysis not only it is difficult to construct high exacted common pronunciation recognize but also to improve the accuracy of it owing to the limitation of spectrum characteristics. We were told that if the sampling is come many people in the RTSR, the accuracyy only reaches 95% to recognize 30 words. The less the sampling, the lower the accuracy.

It is apparent that spectrum analysis is not only way of pronunciation recognition. Otherwise, it is difficult to explain whether man-brain can recognize different person's speaking and why a child lived in female society for a long time can hear and understand the men's talking.

In fact, human pronunciation is a typical fuzzy phenomena . Because man-brain has capability of bandling fuzzy information, its capability recognizing pronunciation is greater than one of spectrum recognizer without considering fuzzy influence.

The fuzziness of pronunciation phonomena mostly come from that people enunciate differently. The enunciation of a word hasn't unit rule in the beginning and end, leant and shortage,frequency, strongness and weakness. Spectrum analysis is mainly a analysis from the angle of the frequency combination. It is needed to suppose that the pronunciation of word is conformity with the united model of same

frequency combination. For a person, it may be true under a certain condition,but there are problems for common to this supposition.

Considering generally, fuzzy transition information can be used and new way of recognition may be found. How is the transitivity inside pronunciation analyzed? A simple method is to diffuse information gainned every time . The measurement of this information is defined in the amplitude of pronunciation in researched time . If so, information diffusing is pronunciation vibrating diffusing. The piling up result of information diffusing every time is stand for a vibration combination. Owing to considering the fuzziness of pronunciation recognition , it has a common signition. The cause which principle of information diffusing can be used is that if the time $t_1$ is adequate close with time $t_2$ , the pronunciation information $y_1$ fallen in point $t_1$ is almost equal to the pronunciation information $y_2$ fallen in point $t_2$. But when pronunciation is expressed by vibration loads, the amplitude records by physic equipment is changed rapidly. From vibration envelope, you can see the essence of speech sound that there is fuzzy information appearing by transition. But if we use envelope to recognize, we ignor the combination characteristics of vibration. Using the principle of information diffusing for pronunciation recognition, we can use the information of both, and get a better model .

Because we are interested in the events in whole pronunciation records time interval, the piling up results of inference each other of information in this time interval is significant. So the information diffusing range in pronunciation recognition is only finite and it's the maximum can't surpass the length of whole record. Owing to the length of word or phrase pronunciation is finite, the length of record is also so and diffusing coefficient is not 0.

The technique of injecting information from central point and diffusing it layer by layer is a simple method of pronunciation recognition using principle information diffusing. The principle of this method is that injecting information each time is diffused in records to take them piling up and standardizing and recognition model is established.

## 6. RECOGNITION PRONUNCIATION WITH THE TECHNIQUE OF FALLEN INFORMATION FORM A CENTRAL POINT AND DIFFUSING IT LAYER BY LAYER.

Four phrases are analyzed . We have established the model with male's pronunciation for recognizing the female's voice.

The four phrases are "provincial academy of science", "provincial socioacademy of science", "provincial agriculture academy of science", and "horticulture institute" . The male's pronunciations of the four phrases are signed $M_1, M_2, M_3, M_4$ and the female's pronunciations are signed $W_1, W_2, W_3, W_4$ respectively. We can record these pronunciations with general cassette recorder.Using module shifter,we can exchange it into digit-information with the sampling frequency in 400 points per second. Because the vibration record are symmetrical roughly, we can only use the half branch up record to analysis. In four phrases, let's cancel the obvious pause between words and part of nonpronunciation record at the head and the tail and obtain the pronunciation digitizing crude information as shown in appendix 1.

Because the long and short, strong and weak of varied phrases is not same, it's necessary to have a standardization processing before analysis. We consider

the square root of mean value of amplitude–$\bar{M}_i, \bar{W}_i$ $(i = 1,2,3,4)$ as strong and weak elements. Using the method of information distribution between two points in equi-interval way at appendix 1, and then the truth record amplitude is handled with information distribution between two points. Let's normalize the information in appendix 1 and it's results shows in appendix 2.

In the light of following processing, we will be disposed of pronuncial digital information in appendix 2 with the technique of fallen information from a central point and diffusing it layer by layer in the whole universe. As a result of record length being 300 points, based on diffusing model of Fig.3, the diffusing coefficient is

$$C_{300} = (c_1, c_2, \cdots, c_{300}) = (1.698 \times 10^{-2}, 1.696 \times 10^{-2}, \cdots, 1.683 \times 10^{-38}) \quad (7)$$

Let $A(t_j)$ be the amplitude of pronunciation record A at t and $Q(t_j, t_k)$ is the result of that information fallen from point $t_j$ diffuse to point $t_k, Q(t_j, t_k)$ is calculated by the fallowing formula:

$$Q(t_j, T_k) = C_i A(t_j), \quad i = |j - k| + 1 \quad (8)$$

Now let's give an example in appendix 2. If $M_1(t_5) = 1.895 \times 10^{-2}$, , then

$$\begin{aligned} &Q(t_5, t_5) = 1.698 \times 10^{-2} \times 1.859 \times 10^{-2} = 3.157 \times 10^{-4} \\ &Q(t_5, t_6) = Q(t_5, t_4) = 1.696 \times 10^{-2} \times 1.859 \times 10^{-2} = 3.153 \times 10^{-4} \\ &\cdots\cdots\cdots\cdots \\ &Q(t_5, t_{300}) = C_{296} M_1(t_5) = 3.649 \times 10^{-38} \times 1.859 \times 10^{-2} = 6.783 \times 10^{-40} \end{aligned} \quad (9)$$

Piled up the information of all time points and divided by number of statistical samples, and the result is marked $Q = (q_1, q_2, \cdots, q_{300})$, then

$$q_k = \frac{1}{300} \sum_{j=1}^{300} Q(t_j, t_k), \quad k = 1,2,\cdots,300 \quad (10)$$

Vector Q multiplies the square root of mean value of amplitude A, recognizing model Q are obtained as following:

$$\bar{Q} = (\bar{q}_1, \bar{q}_2, \cdots, \bar{q}_{300}), \quad \bar{q}_k = \bar{A} q_k \quad (11)$$

Disposed of appendix 2, the recognition pattern $\bar{Q}(M_i)$ and $Q(W_i)$ shows in appendix 3.

Of all matching method, here we chose peak response one.The reason is that the peak and its near values is related to numbers of words. The more the word is, the bigger $\bar{M}_i, \bar{W}_i, \bar{Q}(m_i)$ and $\bar{Q}(w_i)$ is. Vice versa .

## REFERENCES

[1] Liu Zhenrong, Application of information distribution concept to the estimation of earthquake intensity , The Analysis of Fuzzy Information, James C.Bezdek, CRC Press, USA, 1987.

[2] Wang Jiading, Huang Chongfu and Liu Zhenrong, Fuzzy information analysis in soil dynamics problem, Fuzzy System and Knowledge Eng. , Guang Dong Higher Education Publishing House, 1987, pp753-760.

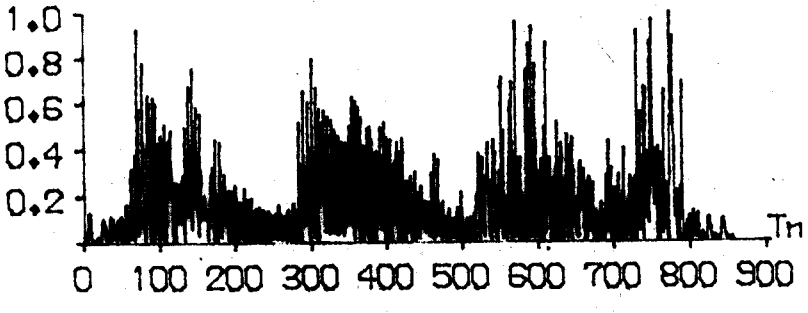
[3] Tao Xiaxin, Fuzzy recognition of seismic pattern, Fuzzy System and Knowledge Eng., Guang dong Higher Education Publishing House, 1987, pp531-532.

[4] Liu Zhenrong and Huang Chongfu, Information distribution method relevant in fuzzy information analysis,Fuzzy sets and Systems, 36 (1990) pp.67-76.

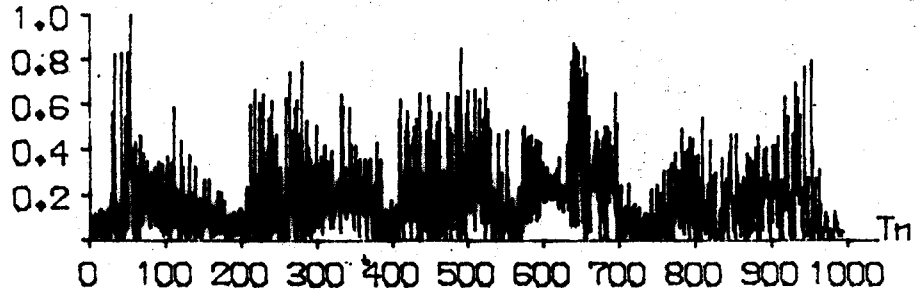
[5] C. S.Myers and L.R.Rabiner,Connected digited recognition using a level-building DTW algorithm, IEEE Trans., June 1981, ASSP-21 351.

APPENDIX 1

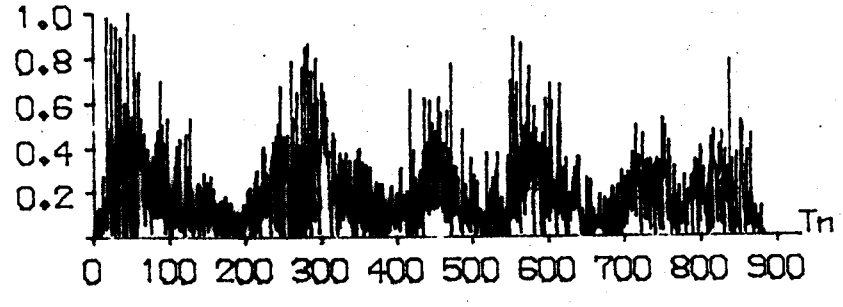
(1) The digital information if male voice



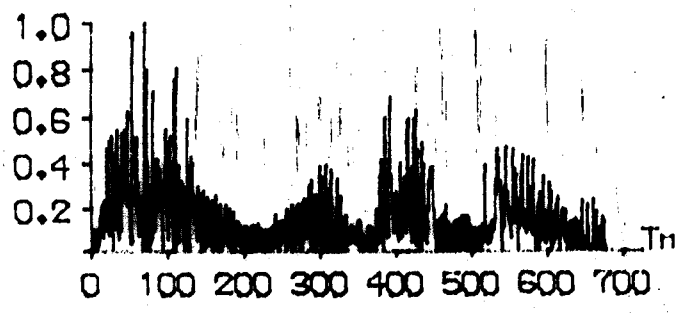

$M_1$ = provincial academy of science, $\bar{M}_1 = 0.1626$




$M_2$ = provincial socioacademy of science, $\bar{M}_2 = 0.1838$




$M_3$ = provincial agriculture academy of science, $\bar{M}_3 = 0.1932$




$M_4$ = horticulture institute, $\bar{M}_4 = 0.1469$

(2) The digital informationif female voice

(omit)

APPENDIX 2-3 ( omit )