ISSN 2278-3091
Volume 6, No.3, May - June 2017
International Journal of Advanced Trends in Computer Science and Engineering
Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse04632017.pdf

The distribution formation method of reference patterns of vocal speech sounds

E. Fedorov[1], H. Alrababah[2], A. Nehad[3]

[1] Professor, Ukraine, fedorovee75@donntu.edu.ua

[2] Dr., UAE, hamza@uoj.ac.ae

[3] Mr., UAE, ali@uoj.ac.ae

ABSTRACT

The development of intelligent computer components is widespread problem. The method of distributed transformation training patterns vocal sounds to a unified amplitude-time window and method a distributed clustering training patterns vocal sounds have been proposed for distributed forming reference patterns of speech vocal sounds in the paper. These methods allow fast convert quasi-periodic sections of different lengths to a single amplitude-time window for subsequent comparison and accurately and quickly determine the optimal number of clusters, which increases the probability clusterization. The proposed methods can be used in speech recognition and synthesis systems.

Keywords: distributed transformation training patterns vocal sounds, unified amplitude-time window, distributed clustering training patterns vocal sounds.

1. INTRODUCTION

The general formulation of the problem. The development of intelligent software components intended for human speech recognition, speech synthesis et al., which are used in computer systems to communicate, is actual in current conditions. The basis of this problem lies the problem of building effects in efficiently methods that provide high speed formation of the reference patterns of speech sounds and used to learn descriptive and generative models.

Analysis Research. Existing speech patterns synthesis system using such approaches like [1-4]: formant synthesis, synthesis based of a linear prediction coefficients (LPC-synthesis), concatenative synthesis. Formant synthesis and LPC-synthesis are based on the model of human speech formation. The model of the speech path is realized as an adaptive digital filter. For formant synthesis parameters of the adaptive digital filter are determined by the formant frequencies [5, 6], and LPC-synthesis - LPC coefficient [7]. The best results regarding the intelligibility and naturalness of the sound of speech can be obtained by concatenative synthesis. Concatenative synthesis is carried out by gluing the necessary sound units [1,3,8,9]. In such systems, it is necessary to apply signal processing to bring the frequency of the pitch, energy and duration of the sound units to those with which the synthesized speech should be characterized. In the systems of concatenative synthesis, three main algorithms are used: TD-PSOLA (made scaling the audio unit of time), FD-PSOLA (made scaling the audio unit of frequency), LP-PSOLA (carried scaling the prediction error signal in time with the subsequent application of LPC filter-coefficients). The disadvantage of concatenative synthesis is the need to store a large number of sound units. In this connection, the problem arises of their more economical representation [10]. Existing speech pattern recognition systems utilize approaches such as: logical, metric, Bayesian, connectionist, generative. Modern methods and speech pattern recognition models are usually based on: Hidden Markov models [5], CDP (Composition + Dynamic Programming) [10], Artificial neural networks [11 - 15] and have the following disadvantages [16, 17]: unsatisfactory probability of recognition; the need for a large number of training data; duration of training; storage of a large number of reference of sounds or words, as well as weight coefficients; duration of recognition.

Formulation of research problems. The aim is to develop a method of forming a distributed reference patterns vocal speech sounds.

Problem Solving and research results. To achieve this aim you need:

1. Develop a distributed method conversion training patterns vocal sounds to a unified amplitude-time window.
2. Develop a method for distributed clustering training patterns vocal sounds.
3. Conduct a numerical study of the clustering methods used.

2. METHOD OF CONVERTING A DISTRIBUTED TRAINING PATTERNS VOCAL SOUNDS TO A UNIFIED AMPLITUDE-TIME WINDOW

Let defined finite set of training patterns vocal sound which is described by a set of limited finite integer discrete functions $X=\{x_i \mid i\in\{1,...,I\}\}$, where $A_i^{\min}, A_i^{\max}$ –

minimum and maximum value of the function x_i on a compact. $\{N_i^{\min},..., N_i^{\max}\}$. We introduce the following mean values

$$\breve{N}^{av} = \frac{1}{I}\sum_{i=1}^{I}\left(N_i^{\max} - N_i^{\max}\right),$$

$$\breve{A}^{av} = \frac{1}{I}\sum_{i=1}^{I}\left(A_i^{\max} - A_i^{\min}\right).$$

Let I - the initial number of parallel threads. Let first the thread number correspond to the number of the training pattern of vocal sound. Then each i-th thread by the transformation described in [18, 19] maps the function x_i into an integral bounded finite discrete function s_i, and the function s_i has compact support $\{1,..., \breve{N}^{av}+1\}$ and minimal value 0 and maximal value $\breve{A}^{av}$ on it. As a result of all the threads will be received family $S = \{s\}$.Thus, it is possible to quickly convert quasi-periodic signal portion of different lengths to a unified amplitude-time window for subsequent comparison.

3. THE METHOD OF DISTRIBUTED CLUSTERING TRAINING PATTERNS OF VOCAL SOUNDS

Suppose now that the thread number corresponds to the number of clusters into which the family S. Then each thread with the number K performs clustering method with the number of clusters K. As a result of the work for each K-th thread the pair $(J_K, \{m_{Kk} \mid k \in \{1,...,K\}\})$ will be received, where J_K - the value of the objective function, m_{Kk} - cluster center k. Choosing the best pair numbers of all threads is performed in the form

$$K^* = \min\{K \mid J_K < \delta, K \in \{1,..., I\}\}$$

As a result, many reference patterns will be created

$$H = \{h_k \mid h_k = m_{K^*k}, k \in \{1,...,K^*\}\}.$$

Thus, it becomes possible to accurately and quickly determine the optimal number of clusters. Researched article iterative clustering methods are shown below

2.1. Clustering Method based on an algorithm K-means

1. Initialization:

a) If $K = 1$, then the initial partition $\Re(S) = \{S_1 \mid S_1 = S\}$ on one cluster, which is described by a vector of values of indicator functions $\Lambda = [\chi_{S_1}(s_i)]$, $\chi_{S_1}(s_i) = 1$, $i \in \{1,...,I\}$. Iteration number $\tau = \tau^{\max}$, $\Lambda^* = \Lambda$.

b) If $K = I$, then the initial partition $\Re(S) = \{S_k \mid S_k \subset S\}$ on I clusters, which is described by a matrix of values of indicator functions $\Lambda = [\chi_{S_k}(s_i)]$, $\chi_{S_k}(s_i) = \begin{cases}1, & i = k\\ 0, & i \neq k\end{cases}$, $i \in \{1,...,I\}$, $k \in \{1,.,K\}$. Iteration number $\tau = \tau^{\max}$, $\Lambda^* = \Lambda$.

c) If $1 < K < I$, then the initial partition is given at random $\Re(S) = \{S_k \mid S_k \subset S\}$ on K clusters, which describes initialized randomly matrix of values of indicator functions $\Lambda = [\chi_{S_k}(s_i)]$, $\chi_{S_k}(s_i) = \begin{cases}1, & s_i \in S_k\\ 0, & s_i \notin S_k\end{cases}$, $i \in \{1,...,I\}$, $k \in \{1,.,K\}$. Iteration number $\tau = 1$, $\Lambda^* = O$. In this case, for the matrix Λ the following conditions must be satisfied [20]:

$$\sum_{k=1}^{K}\chi_{S_k}(s_i) = 1,\ i \in \{1,...,I\},$$

$$\sum_{i=1}^{I}\chi_{S_k}(s_i) > 0,\ k \in \{1,.,K\},$$

$$\chi_{S_k}(s_i) \in \{0,1\},\ k \in \{1,.,K\},\ i \in \{1,...,I\}.$$

2. Calculation of cluster centers

$$m_{Kk}(l) = \frac{\sum_{i=1}^{I}\chi_{S_k}(s_i)s_i(l)}{\sum_{i=1}^{I}\chi_{S_k}(s_i)},\ k \in \{1,.,K\},\ l \in \{1,..., \breve{N}^{av}+1\}.$$

3. Distance calculation

$$\forall i \in \{1,...,I\},\ k \in \{1,.,K\}\ \ d_{ik} = \sqrt{\sum_{l=1}^{\breve{N}^{av}+1}|s_i(l) - m_{Kk}(l)|^2}.$$

4. If $1 < K < I$, then modification of the matrix of values of indicator functions is performed according to the following rule

if $k^* = \arg\min_k d_{ik}$, then $\chi_{S_{k^*}}(s_i) = 1$ and $\forall k \in \{1,...,K\}/\{k^*\}$ $\chi_{S_k}(s_i) = 0$.

5. Rule of the termination condition

If $\|\Lambda - \Lambda^*\| > \varepsilon$ and $\tau < \tau^{\max}$, then $\tau = \tau + 1$, $\Lambda^* = \Lambda$, go to 2.

6. The calculation of the objective function

$$k_i^* = \arg\max_k \chi_{S_k}(s_i),$$

$$J_K = \max_i \left(\frac{d_{ik_i^*}}{A\sqrt{\breve{N}^{av} + 1}} \right).$$

2.2. Clustering Method based on an algorithm Fuzzy C-means

1. Initialization:

a) If $K = 1$, then the initial partition $\Re(S) = \{S_1 \mid S_1 = S\}$ on one cluster, which is described by a vector of values of membership functions $M = [\mu_{\tilde{S}_1}(s_i)]$, $\mu_{\tilde{S}_1}(s_i) = 1$, $i \in \{1,...,I\}$. Iteration number $\tau = \tau^{\max}$, $M^* = M$.

b) If $K = I$, then the initial partition $\Re(S) = \{S_k \mid S_k \subset S\}$ on I clusters, which is described by a matrix of values of membership functions $M = [\mu_{\tilde{S}_k}(s_i)]$, $\mu_{\tilde{S}_k}(s_i) = \begin{cases} 1, & i = k \\ 0, & i \neq k \end{cases}$, $i \in \{1,...,I\}$, $k \in \{1,.,K\}$. Iteration number $\tau = \tau^{\max}$, $M^* = M$.

c) If $1 < K < I$, then the initial partition is given at random $\Re(\tilde{S}) = \{\tilde{S}_k \mid \tilde{S}_k \subset \tilde{S}\}$ on K clusters, which describes initialized randomly matrix of values of membership functions $M = [\mu_{\tilde{S}_k}(s_i)]$, where $\mu_{\tilde{S}_k}(s_i)$ return a degree of membership of objects clusters, $i \in \{1,...,I\}$, $k \in \{1,.,K\}$. Iteration number $\tau = 1$, $M^* = O$. In this case, for the matrix M the following conditions must be satisfied [21]:

$$\sum_{k=1}^{K} \mu_{\tilde{S}_k}(s_i) = 1, \; i \in \{1,...,I\},$$

$$\sum_{i=1}^{I} \mu_{\tilde{S}_k}(s_i) > 0, \; k \in \{1,.,K\},$$

$$\mu_{\tilde{S}_k}(s_i) \in [0,1], \; k \in \{1,.,K\}, \; i \in \{1,...,I\}.$$

The weight of fuzzy clustering w is set (in article w =2).

2. Calculation of cluster centers

$$m_{Kk}(l) = \frac{\sum_{i=1}^{I} \left(\mu_{\tilde{S}_k}(s_i)\right)^w s_i(l)}{\sum_{i=1}^{I} \left(\mu_{\tilde{S}_k}(s_i)\right)^w}, \; k \in \{1,.,K\}, \; l \in \{1,..., \breve{N}^{av} + 1\}.$$

3. Distance calculation

$$\forall i \in \{1,...,I\}, \; k \in \{1,.,K\} \;\; d_{ik} = \sqrt{\sum_{l=1}^{\breve{N}^{av}+1} | s_i(l) - m_{Kk}(l) |^2}.$$

4. If $1 < K < I$, then modification of the matrix of values of membership functions is performed according to the following rule

if $d_{ik} > 0$, then $\mu_{\tilde{S}_k}(s_i) = \left(\sum_{l=1}^{I} \left(\frac{d_{ik}}{d_{lk}} \right)^{1/(w-1)} \right)^{-1}$,

if $d_{ik} = 0$, then $\mu_{\tilde{S}_k}(s_i) = 1$ and $\forall k \in \{1,...,K\}/\{k^*\}$ $\mu_{\tilde{S}_k}(s_i) = 0$.

5. Rule of the termination condition

If $\|M - M^*\| > \varepsilon$ and $\tau < \tau^{\max}$, then $\tau = \tau + 1$, $M^* = M$, go to 2.

6. The calculation of the objective function

$$k_i^* = \arg\max_k \mu_{\tilde{S}_k}(s_i),$$

$$J_K = \max_i \left(\frac{d_{ik_i^*}}{A\sqrt{\breve{N}^{av} + 1}} \right).$$

2.3. Clustering method based on EM and -algorithm

1. Initialization:

a) If $K = 1$, then the initial partition $\Re(S) = \{S_1 \mid S_1 = S\}$ on one cluster, which is described by the vector of expected values of hidden variables $G = [g_{i1}]$, $g_{ik} = 1$, $i \in \{1,...,I\}$. Iteration number $\tau = \tau^{\max}$, $G^* = G$.

b) If $K = I$, then the initial partition $\Re(S) = \{S_k \mid S_k \subset S\}$ on I clusters, which is described by the vector of expected values of hidden variables $G = [g_{ik}]$,

$g_{ik} = \begin{cases} 1, & i = k \\ 0, & i \neq k \end{cases}$, $i \in \{1,...,I\}$, $k \in \{1,.,K\}$. Iteration number $\tau = \tau^{\max}$, $G^* = G$.

c) If $1 < K < I$, then the initial partition is given at random $\Re(S) = \{S_k \mid S_k \subset S\}$ on K clusters, which describes initialized randomly matrix of the expected values of hidden variables $G = [g_{ik}]$, $i \in \{1,...,I\}$, $k \in \{1,.,K\}$. Iteration number $\tau = 1$, $G^* = O$. In this case, for the matrix G the following conditions must be satisfied:

$$\sum_{k=1}^{K} g_{ik} = 1,\ i \in \{1,...,I\},$$

$$\sum_{i=1}^{I} g_{ik} > 0,\ k \in \{1,.,K\},$$

$$g_{ik} \in [0,1],\ k \in \{1,.,K\},\ i \in \{1,...,I\}.$$

Each cluster is described by the likelihood function $f(s \mid (m_{Kk}, \sigma^2_{Kk}))$, for which

$$m_{Kk}(j) = \frac{\sum_{i=1}^{I} g_{ik} s_i(j)}{\sum_{i=1}^{I} g_{ik}},\ j \in \{1,..., \breve{N}^{av} + 1\},$$

$$\sigma^2_{Kk}(j) = \frac{\sum_{i=1}^{I} g_{ik} (s_i(j) - m_{Kk}(j))^2}{\sum_{i=1}^{I} g_{ik}},\ j \in \{1,..., \breve{N}^{av} + 1\}.$$

Sets the weighting factor w_k, which w_k corresponds to the a priori probability of appearance of an object from k-th cluster, i.e. $w_k = P((m_{Kk}, \sigma^2_{Kk}))$.

$$w_k = 1/K.$$

2. Calculation of likelihood functions

$$f(s_i \mid (m_{Kk}, \sigma^2_{Kk})) = \frac{1}{\sqrt{(2\pi)^{\breve{N}^{av}+1} \prod_{j=1}^{\breve{N}^{av}+1} \sigma^2_{Kk}(j)}} \cdot$$

$$\cdot \exp\left(-\frac{1}{2} \sum_{j=1}^{\breve{N}^{av}+1} \frac{(s_i(j) - m_{Kk}(j))^2}{\sigma^2_{Kk}(j)} \right),\ k \in \overline{1,K}.$$

3. E-Step (matrix calculation of expected values of hidden variables)

If $1 < K < I$, then calculate g_{ik}, where hidden variable g_{ik} corresponds to the a posteriori probability, i.e. $g_{ik} = P((m_{Kk}, \sigma^2_{Kk}) \mid s_i)$.

$$g_{ik} = \frac{w_{Kk} f(s_i \mid (m_{Kk}, \sigma_{Kk}))}{\sum_{m=1}^{K} w_{Km} f(s_i \mid (m_{Km}, \sigma_{Km}))},$$

4. M-step (calculation parameters w_{Kk}, m_{Kk}, σ^2_{Kk})

$$w_{Kk} = \frac{1}{I} \sum_{i=1}^{I} g_{ik},\ k \in \overline{1,K},$$

$$m_{Kk}(j) = \frac{1}{I w_{Kk}} \sum_{i=1}^{I} g_{ik} s_i(j),\ k \in \overline{1,K},\ j \in \{1,..., \breve{N}^{av} + 1\},$$

$$\sigma^2_{Kk}(j) = \frac{1}{I w_{Kk}} \sum_{i=1}^{I} g_{ik} (s_i(j) - m_{Kk}(j))^2,\quad k \in \overline{1,K},\ j \in \{1,..., \breve{N}^{av} + 1\}.$$

5. Rule of the termination condition

If $\left\| G - G^* \right\| > \varepsilon$ and $\tau < \tau^{\max}$, then $\tau = \tau + 1$, $G^* = G$, go to 2.

6. The calculation of the objective function

$$k_i^* = \arg\max_k g_{ik},$$

$$d_{ik} = \sqrt{\sum_{l=1}^{\breve{N}^{av}+1} | s_i(l) - m_{Kk}(l) |^2},$$

$$J_K = \max_i \left(\frac{d_{ik_i^*}}{A\sqrt{\breve{N}^{av} + 1}} \right).$$

3. NUMERICAL STUDY

Numerical study was conducted for all three clustering methods for 1000 training patterns of vocal sounds of speech. The results are shown in Table 1. According to Table 1, the best results are obtained by EM -algorithm.

Table 1 : Probabilities of clustering

Method name	Probability of clustering, %
K-means	90
Fuzzy C-means	95
EM	98

4. CONCLUSION

The method of distributed transformation training patterns vocal sounds to a unified amplitude-time window and method a distributed clustering training patterns vocal sounds have been proposed for distributed forming reference patterns of speech vocal sounds in the paper. These methods allow fast convert quasi-periodic sections of different lengths to a single amplitude-time window for subsequent comparison and accurately and quickly determine the optimal number of clusters, which increases the probability clusterization. The proposed methods can be used in speech recognition and synthesis systems.

REFERENCES

[1] V.N. Bondarev, F.G. Ade, "**Iskusstvenniy intellect"**, SevNTU, 2002.

[2] R.K. Potapova, "**Rech: kommunikatsia, informatsia, kibernetika**", Radio i sviaz, 1997.

[3] T. Dutoit, "**An introduction to text-to-speech synthesis**", Kluwer Academic Publishers, 1997.

[4] J. Allen, S. Hunnicut, D. Klatt, "**From text to speech, the MITALK system**", Cambridge University Press, 1987.

[5] L.R. Rabiner, R.V. Shafer, "**Digital Processing of Speech Signals**", Prentice-Hall Inc., 1978

[6] G. Bailly, G. Murillo, O. Dakkak, B. Guerin, (1988), "**A text-to-speech system for French using formant synthesis**", Proc. of SPEECH' 88, pp. 255-260.

[7] L.R. Rabiner, B.H. Jang, "**Fundamentals of speech recognition**", Prentice Hall PTR, 1993.

[8] A.J. Hunt, A. Black, "**Unit selection in a concatenative speech synthesis system using a large speech database**", ICASSP 96, pp. 11-14.

[9] C. Hamon, E. Moulines, F. Charpentier, "**A diphone system based on time-domain prosodic modifications of speech**", Proc. of ICASP 89, pp. 238-241.

[10] T.K. Vintsiuk, "**Analiz, raspoznavanie i interpretatsia rechevikh signalov**", Naukova dumka, 1987.

[11] S.O. Haykin, "**Neural Networks and Learning Machines**", Pearson Education, Inc., 2009.

[12] T. Kohonen, "**Self-organizing Maps**", Springer-Verlag, 1995.

[13] R. Callan, "**The essence of neural networks**", Prentice Hall Europe, 1998.

[14] S.N. Sivanandam, S. Sumathi, S.N. Deepa, "**Introduction to Neural Networks using Matlab 6.0**", The McGraw-Hill Comp., Inc., 2006.

[15] K.-L. Du, M.N.S. Swamy, "**Neural Networks and Statistical Learning**", Sprnger-Verlag, 2014.

[16] E.E. Fedorov, **"Iskusstvennyie neyronnyie seti**", DVNZ «DonNTU», 2016.

[17] E.E. Fedorov, **"Metodologia sozdania multiagentnoi sistemy rechevogo upravlenia**", Noulidzh, 2011.

[18] E.E. Fedorov, "**Metod klassifikatsii vokalnyih zvukov rechi na osnove saundletnoy bayesovskoy neyronnoy seti**", Upravlyayuschie sistemyi i mashinyi, Vol. 6, pp. 78-83.

[19] E.E. Fedorov, "**Metod sinteza vokalnyih zvukov rechi po etalonnyim obraztsam na osnove saundletov**", Naukovi pratsi Donetskogo natsionalnogo tehnichnogo universitetu, Vol. 2, pp. 127-137.

[20] K. Ahuja1, A. Sain, "**Analyzing formation of K Mean Clusters using similarity and dissimilarity measures**", International Journal of Advanced Trends in Computer Science and Engineering, Vol. 2 , No.1, pp. 72-74.

[21] S. Baboo, S. Priya, "**Clustering based integration of personal information using Weighted Fuzzy Local Information C-Means Algorithm**", International Journal of Advanced Trends in Computer Science and Engineering, Vol. 2, No.2.