Automatic Speech Evaluation and Multi-model Feedback Language Training to Articulation Disorders

Yeou-Jiunn Chen[a*], Jiunn-Liang Wu[b], Hui-Mei Yang[b]
[a]Department of Electrical Engineering, Southern Taiwan University,
#1, Nan-Tai Street, Yung-Kung City, Tainan County, Taiwan
[b]Department of Otolaryngology, National Cheng Kung University Hospital,
#138, Sheng Li Road, Tainan, Taiwan
*Corresponding Author: chenyj@mail.stut.edu.tw

ABSTRACT

Articulation errors will seriously reduce speech intelligibility and the ease of spoken communication. Typically, a speech-language pathologist uses his or her clinical experience to identify articulation error patterns, a time-consuming and expensive process. Moreover, the language training is very difficult to articulation disorders for personal training procedure. In this paper, a system with automatic speech evaluation and multi-model feedback language training is proposed to assist speech-language pathologists and articulation disorders. The articulation error patterns in phonetic can be identified and considered to generate pronunciation confusion network. Using speech recognition technique, the pronunciation errors can be automatically identified and labeled by dependency network. The articulation error pattern can be identified by dependency network. For language training, multi-model feedback interface is also proposed to assist articulation disorders. A 3D virtual facial animation with speech signal, lip motion, and tongue motion is proposed to promote users to articulate clearly. Moreover, the articulation teaching activities for each type of articulation errors is also feed back to caregivers. It can promote articulation disorders to improve articulatory abilities in games. Experimental results reveal the practicability of proposed method and system.

Keyword: Articulation Disorder, Articulation Error Pattern, Multi-model Feedback, 3D Virtual facial animation, Dependency Network

1. Introduction

Articulation errors, which generate different degrees of abnormality in articulation, seriously reduce speech intelligibility and the ease of spoken communication. Typically, a speech-language pathologist uses his or her clinical experience to identify articulation error patterns, a time-consuming and expensive process. Moreover, the

language training is very difficult to articulation disorders for personal training procedure. Therefore, an automatic process for identification of articulation error patterns and personal language training are very helpful to assist speech-language pathologists and articulation disorders.

Most articulation errors fall into those three categories: omissions, substitutions, or distortions. For speech-language pathologist, the articulation errors are examined in terms of the place and manner of articulation and can be classified into six articulation error patterns: fronting, backing, de-aspiration, stopping, affrication, and omission (Bernthal et al., 2004). In a typical fronting error, for example, a child may say /t/ instead of /k/ in the Chinese word /kan4/ so it would be heard as /tan4/. For backing error, the /q/ will be pronounced as the /k/ and /qi4/ would be heard as /ki4/.

Recently, researchers proposed various approaches to identify articulation errors using statistical models (Georgoulas et al., 2006) or tongue detection models (Sterns et al., 2006, Hueber et al., 2007, Robineau et al., 2007). For statistical models, Georgoulas et al. applied support vector machine to classify only three consonant phonemes. Only those phonemes were insufficient to identify the articulation error patterns and the error information of pronunciation was insufficient. For tongue detection models, using ultrasound to examine speech production was gaining popularity because of its portability and noninvasiveness. The place of tongue could be detected from the ultrasound image. However, the resolution of detected results was insufficient to distinguish the five articulation error patterns. Moreover, the manner of articulation was also cannot be detected by this approaches.

Automatic speech recognition (ASR) with spontaneous speech had been widely applied to many applications (Ali et al., 2001, Siniscalchi et al., 2007, Rajamanohar et al., 2005, Orzechowski et al. 2005, Li et al., 2005, Ali et al., 2001b, Mak et al., 2003). The features used in ASR were succeeded to identify the articulation attributes, which correlate closely with the place of articulation and the manner of articulation (Siniscalchi et al., 2007, Rajamanohar et al., 2005, Orzechowski et al. 2005, Li et al., 2005, Ali et al., 2001b, Mak et al., 2003). Hence, ASR will be very useful to identify the pronunciation errors without manually labeling the articulatory information. Testing vocabularies with multiple syllables can provide more articulatory information and reduce the evaluation time in clinical speech evaluation.

Besides, dependency network (DN) had been applied to collective classification and knowledge discovery (Tian et al., 2006, Preisach et al., 2006). It is well suited to the task of predicting preferences and is generally useful for probabilistic inference. From the clinical practice and experience, the graphical representation of DN can be easily designed to represent the relationships of speech, testing vocabulary, and articulation error pattern. Thus, it is appropriate to integrate clinical experience into identification

of articulation error patterns.

For language training, researches usually aimed at computer-assisted treatment by using visual feedback (Sheng et al., 2001). However, visual feedback was focus on speech parameters, such as spectrogram, pitch contour, and formants. For secondary language learning, the lip animation and sounds can be responded to users. However, the main factors for articulation disorder are the wrong movement of tongue. There are lots of previous works in early periods, and most of them focused on 3D tongue and lip models, which were generated from electromagnetic articulography, ultrasound, electropalatography, cineradiography, x-ray micro-beam, and magnetic resonance imaging (Westbury et al., 1994, Kramer et al., 1991, Munhall et al., 1995, Le Goff, 1997). By those approaches the movement could be estimated and used to generate the 3D tongue and lip animation. However, the 3D images are very difficult to be acquired and the generation of 3D tongue is also very complex.

In this paper, a novel automatic approach integrating automatic speech recognition, dependency network, 3D facial animation, and language training activities is proposed to assist speech-language pathologists and articulation disorders. First, the types of articulation error patterns are determined by automatic speech recognition and dependent network. Secondary, the text corpus for language training can be selected. The 3D facial animation with speech signal, lip movement, and tongue movement is applied to help user in simulation of articulatory behaviors. Finally, the training strategies in games are designed in multimedia format and provided to caregivers.

2. Methods

The system architecture of automatic speech evaluation and multi-model feedback language training is shown in Fig. 1. First, a photo naming task (PNT) is used to capture examples of an individual's articulation patterns. Using ASR and dependency network, the articulation error patterns can be effectively identified. Then, 3D facial animation including speech signal, lip movement, and tongue movement is applied to promote user to articulate clearly. Moreover, a suitable articulation training activities in games is also provided to caregiver and improve articulatory ability in games.

2.1 Photo naming task

In clinical protocol, speech-language pathologists use PNT represented as pictures to obtain the articulatory information of a child in speeches. The articulation error patterns of a articulation disorder can be identified from those speeches. Therefore, PNT should include familiar vocabulary words with recognizable pictures. This will decrease or eliminate the need for the child to imitate the clinician when presenting test stimulus items. Second, it should assess the production of all phonemes in a

specific language. Those phonemes should be presented in at least two different word positions. Third, it should assess sounds in increasingly complex contexts. It should include target sounds in mono-syllabic and multi-syllabic words. Consequently, PNT with a set of familiar words can be written as

$$W=\{w_1, w_2, \ldots, w_N\} \quad (1)$$

where N is the number of testing words. Each testing word w_i can be treated as a concatenation of phonemes and written as

$$w_i = s_1 s_2 \cdots s_{N_i} \quad (2)$$

where s_j is the *j-th* phoneme and N_i is the number of phonemes in w_i. Therefore, PNT can be treated as a set of phonemes and written as

$$S=\{s_1, s_2, \ldots, s_M\} \quad (3)$$

where M is the number of phonemes appeared in each testing words of W. Each s_j in S is modeled as Hidden Markov Model (Sher et al., 2006).

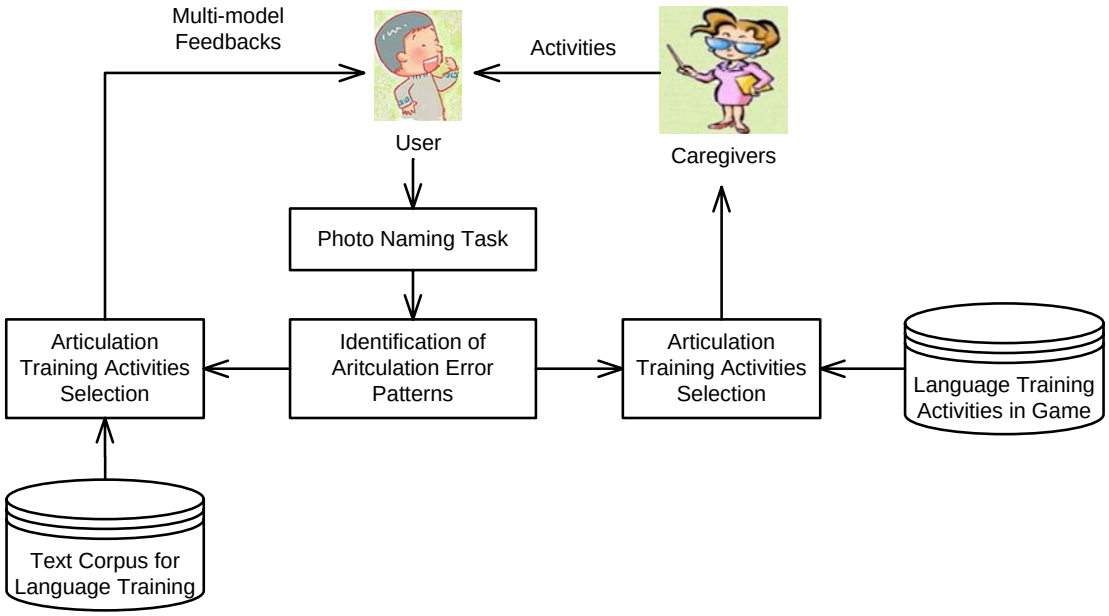

Fig. 1. System architecture of automatic speech evaluation and multi-model feedback language training

2.2 Identification of articulation error patterns

In the process of clinical speech evaluation, a testing subject is actuated to articulate s_m and the corresponding speech observation o_m can be acquired. With o_m and s_m, a speech-language pathologist have to manually label the speech observation o_m as $\hat{s}_m$. Therefore, a dependence network for labeling process is constructed and shown in Fig. 2a. The joint probability of dependency network consists of a set of conditional probability distributions:

$$P\left(\hat{s}_m s_m o_m\right) = P\left(\hat{s}_m \left| s_m o_m \right.\right) P\left(o_m\right) P\left(s_m\right). \tag{4}$$

Since, the labeling process of $\hat{s}_m$ comprises linguistic and acoustic information, the node $\hat{s}_m$ is designed as the combination of $\hat{s}_m^l$ and $\hat{s}_m^a$. The dependency network for automatic labeling process is modified and shown in Fig. 2b. $\hat{s}_m^l$ and $\hat{s}_m^a$ are the labeling results based on linguistic and acoustic information, respectively. Therefore, the conditional probability of node $\hat{s}_m$ can be written as

$$P\left(\hat{s}_m \left| s_m o_m \right.\right) \cong \left(P\left(\hat{s}_m^l \left| s_m \right.\right)^{\omega_l} P\left(\hat{s}_m^a \left| o_m \right.\right)^{\omega_a} \right)^{1/\omega_l + \omega_a}. \tag{5}$$

w_l and w_a are the weighting factors for language and acoustic information. $P\left(\hat{s}_m^l \left| s_m \right.\right)$ and $P\left(\hat{s}_m^a \left| o_m \right.\right)$ are the probabilities of language and acoustic.

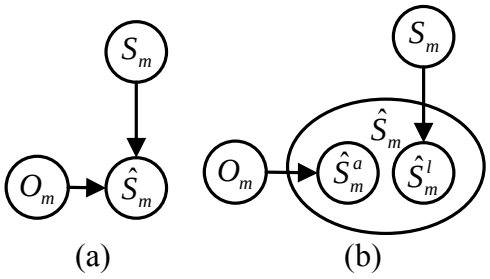

Fig. 2 Dependency network for labeling process of s_m (a) original (b) modified for automatic labeling

A speech-language pathologist uses target phoneme s_m and labeled phoneme $\hat{s}_m$ to identify the i-th articulation error pattern E_m^i. Hence, the dependency network for identification of articulation error pattern is designed and shown in Fig. 3. For a phoneme s_m, the probability of articulation error pattern can be estimated as

$$\begin{aligned} P\left(E_m^i s_m \hat{s}_m o_m\right) &= P\left(E_m^i \left| s_m \hat{s}_m \right.\right) P\left(\hat{s}_m \left| s_m o_m \right.\right) P\left(o_m\right) P\left(s_m\right) \\ &\approx P\left(E_m^i \left| s_m \hat{s}_m \right.\right) \left(P\left(\hat{s}_m^l \left| s_m \right.\right)^{\omega_l} P\left(\hat{s}_m^a \left| o_m \right.\right)^{\omega_a} \right)^{1/\omega_l + \omega_a} P\left(o_m\right) P\left(s_m\right) \end{aligned} \tag{6}$$

When a testing subject is actuated to articulate $w_i = s_1 s_2 \cdots s_{N_i}$, the corresponding speech observations O_i can be recorded. A DN defined in Fig. 3 is applied to automatically find the segmentations $O_i = o_1 o_2 \cdots o_{N_i}$ and corresponding labeling

results $\hat{w}_i = \hat{s}_1 \hat{s}_2 \cdots \hat{s}_{N_i}$ with maximum posterior probability as follows:

$$\begin{aligned} \hat{w}_i &= \arg\max_{\tilde{w}_i} P\left(\tilde{w}_i w_i O_i\right) \\ &= \arg\max_{\tilde{s}_1 \tilde{s}_2 \cdots \tilde{s}_{N_i}} P\left(\tilde{s}_1 s_1 o_1 \tilde{s}_2 s_2 o_2 \cdots \tilde{s}_{N_i} s_{N_i} o_{N_i}\right) \end{aligned} \tag{7}$$

For articulation disorders, the articulatory information of each phoneme are consistent and the coarticulation effect is modeled by context-dependent models. Thus, each phoneme is assumed to be independent and Eq. (7) integrated with Eq. (4) and Eq. (5) can be derived as

$$\begin{aligned} \hat{w}_i &= \arg\max_{\tilde{s}_1 \tilde{s}_2 \cdots \tilde{s}_{N_i}} \prod_{j=1}^{N_i} P\left(\tilde{s}_j s_j o_j\right) \\ &= \arg\max_{\tilde{s}_1 \tilde{s}_2 \cdots \tilde{s}_{N_i}} \prod_{j=1}^{N_i} P\left(\tilde{s}_j \middle| s_j o_j\right) P\left(o_j\right) P\left(s_j\right) \\ &= \arg\max_{\tilde{s}_1 \tilde{s}_2 \cdots \tilde{s}_{N_i}} \prod_{j=1}^{N_i} \left(P\left(\tilde{s}_j^l \middle| s_j\right)^{\omega_l} P\left(\tilde{s}_j^a \middle| o_j\right)^{\omega_a} \right)^{1/\omega_l + \omega_a} P\left(o_j\right) P\left(s_j\right) \end{aligned} \tag{8}$$

As the observation o_j and target s_j are constant across each estimation, the denominators $P(o_j)$ and $P(s_j)$ are omitted to reduce the complexity of estimation. Finally, the probability of labeling and segmentation can be estimated as

$$\hat{w}_i = \arg\max_{\tilde{s}_1 \tilde{s}_2 \cdots \tilde{s}_{N_i}} \prod_{j=1}^{N_i} \left(P\left(\tilde{s}_j^l \middle| s_j\right)^{\omega_l} P\left(\tilde{s}_j^a \middle| o_j\right)^{\omega_a} \right)^{1/\omega_l + \omega_a} \tag{9}$$

In this paper, $P\left(\tilde{s}_j^a \middle| o_j\right)$ is estimated by HMM and $P\left(\tilde{s}_j^l \middle| s_j\right)$ is estimated by maximum likelihood estimation (MLE).

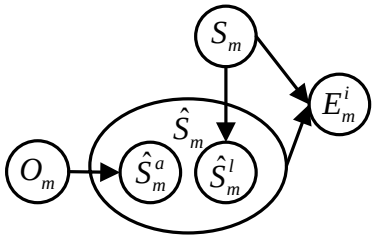

Fig. 3 Dependency network to identify articulation error pattern E_m^i with s_m

2.3 Multi-model feedback language training

A 3D facial animation including speech signal, lip movement, tongue movement is proposed and shown in Fig. 4. Text-to-speech is integrated to generate speech signal

and corresponding syllable boundaries. For phoneme segmentation, the contextual knowledge represented as consonant (C) and vowel (V) sequence is used to select boundary detection criterions in sequential forward selection. Let $O = o_i \in R^d$, $i=1,2,\ldots,N$, be a sequence of framed-based cepstral vectors extracted from a speech signal. It is important to find a boundary at frame $b \in (1, N)$. Contextual knowledge based phoneme segmentation is applied to identify the phoneme boundaries of input text. First, the Hotelling's T^2 test is adopted to identify the CV (C following V or V following C) boundaries. Using likelihood-ratio procedure approach, the Hotelling's T^2 test statistic can be written as

$$T_b^2 = y_b' \Sigma_b^{-1} y_b \tag{10}$$

where

$$y_b = \sqrt{\frac{b(N-b)}{N}} (\mu_1 - \mu_2). \tag{11}$$

μ_1 and μ_2 is the mean of speech segments before and after boundary b, respectively. Σ is the common covariance matrix of O. For VV (V following V) boundaries, Bayesian information criterion is then applied to measure the difference of boundary b and written as

$$\Delta BIC(b) = \frac{1}{2}\left(N \log|\Sigma| - b \log|\Sigma_1| - (N-b)\log|\Sigma_2|\right) - \frac{1}{2}\lambda\left(d + \frac{1}{2}d(d+1)\right)\log N \tag{12}$$

where Σ_1 and Σ_2 are the variance of segments before and after boundary b. λ is the penalty factor to compensate for small sample size cases.

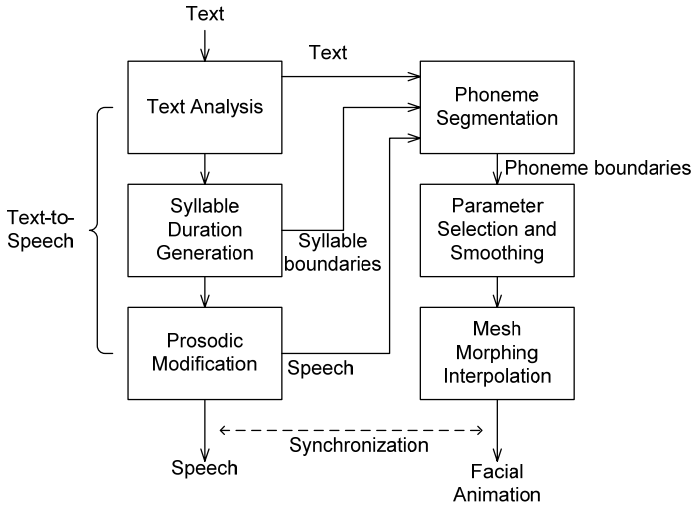

Fig. 4. The block diagram of 3D facial animation

For mandarin speech, there are 16 vowels and 21 consonants. Base on knowledge of acoustic phonetics, there are 105 categories of lip motions can be defined to represent the lip motions of all 408 Mandarin syllables. Given a feature point with location (x_t, y_t) in frame t, the location of this feature point in frame t+1 is decided by:

$$(x_{t+1}, y_{t+1}) = \arg\min_{(x_{t+u}, y_{t+v})} \left(\sum_{i=\frac{-m}{2}}^{\frac{m}{2}} \sum_{j=\frac{-m}{2}}^{\frac{m}{2}} \left(I(x_t + i, y_t + j) - I(x_t + i + u, y_t + j + v)^2 \right) \right) \quad (13)$$

where $-\frac{z}{2} \le u \le \frac{z}{2}$ and $-\frac{z}{2} \le v \le \frac{z}{2}$. $I(x, y)$ indicates the intensity of pixel (x, y). z is the block size that contains the possible location of the feature point in frame $t+1$. Finally, the control points are transformed to feature points of 3D facial models. Tongue models had been used in many research areas and non-offensive estimation was proposed in this paper. The 2D animations of mouth cavity for articulatory are collected. For each mouth cavity pictures of 2D animation of a phoneme, sobel operator is applied to detect the edge of tongue. Base on the edge of tongue, the turning points are selected as feature points and mapping with the parametric points of 3D tongue models. Fig. 5 shows an example of detected control points and corresponding parametric points of 3D tongue models. Finally, the B-spline is applied to smooth the sequence of parametric points of 3D facial models.

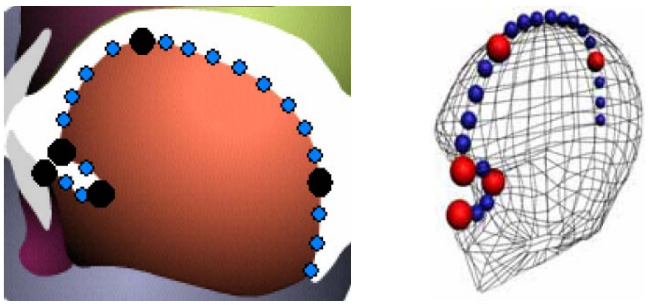

Fig. 5. An example of detected control points and the mapping feature points of tongue models

2.4 Articulation training strategies in games

In the clinical protocol, the sequence of treatments for articulation disorder is: (1) establishment of self articulation problem consciousness; (2) discriminative of correct or incorrect articulation; (3) establishment of correct articulation; (4) language therapy/training in games (Yu, 1992, Wu, 2000). Therefore, five training strategies named articulation error awareness, speech discrimination training, provoking articulation practice, articulation habit treatment, and sentence based articulation training are designed for each type of articulation errors. The results of language

training activities are shown in Table I.

The relationship of articulation error types between each phoneme is analyzed by language therapists and shown in Table I. The number in Table I is the no of articulation errors. Integrating Table I and detected articulation errors, the probability for each type of articulation errors can be estimated. Using statistical process, a threshold can be objectively selected by language therapist and used to eliminate some incorrect decisions generated by speech recognition errors. According to articulation characteristics, the language training activities in games are designed by language therapists and used help articulation disorders to correct articulation habit. Some language training activities in games are shown in Table II.

Table I. The relationship of articulation error types between each phoneme

		Pronounced					
		b	p	m	f	d	t
Target	b						
	p	3					
	m	4	4			4	4
	f	4	4			4	4
	d						

Table II. Part of language training activities in games

No	Language training activities
1	Utilize action of the cough, child is taught what the throat sound is.
2	Utilize the action of blow, child is taught what aspiration is.
3	Utilize the enunciation simulation, child is taught what alveolar is.
4	Pronounce the long sound x~~yu ,It is fricative.

3. Experimental Results

3.1 Results of identifying articulation error patterns

Samples were collected from 553 children (346 males and 207 females) with multiple articulation error patterns. 421 and 132 samples were used for training and testing, respectively. The articulation error patterns of those samples were manually labeled by speech-language pathologists. In the training database, there are 45, 179, 88, 297, 106, and 42 samples for fronting, backing, de-aspiration, stopping, affrication, and omission, respectively. Moreover, in the testing database, there were 15, 57, 28, 95, 33, and 13 samples for fronting, backing, de-aspiration, stopping, affrication, and omission, respectively.

The priori probability of each pronunciation error, $P\left(E_m^i \left| s_m \hat{s}_m \right.\right)$, is different to determine the articulation error patterns and should be estimated in the training database. The probability of distributions of pronunciation errors for articulation error patterns is shown in Table III. It is clear that the correlation between pronunciation error and articulation error pattern is different. Some pronunciation errors can give confident to identify an articulation error pattern. However, for clinical practice, to identify an articulation error pattern should be verify by different phoneme's pronunciation characteristic.

Table III. Probability distributions of pronunciation error for each articulation error pattern

Fronting		Backing		De-aspiration	
PE	PB	PE	PB	PE	PB
g→d	83.33	d→g	87.96	p→b	73.99
k→d	76.92	t→k	92.57	t→d	68.79
k→t	81.08	zi→d	16.94	k→g	7.41
zhi→d	18.80	zi→g	93.09	q→j	65.96
chi→d	27.48	ci→d	10.42	ci→zi	76.92
chi→t	15.85	ci→t	14.89		
Stopping		**Affrication**		**Omission**	
PE	PB	PE	PB	PE	PB
f→b	75.26	x→j	66.15	b→NULL	87.50
l→g	90.91	shi→zi	81.58	p→NULL	45.61
ri→g	83.87	si→zi	70.97	m→NULL	38.89
zi→d	28.13			f→NULL	47.50
zi→g	96.28			d→NULL	71.79
ci→d	89.58			t→NULL	75.00
ci→t	87.23			n→NULL	23.26

PE: Pronunciation Error
PB: Probability

To decide the identification results, a threshold of articulation error pattern should be determined. The receiver operating characteristic (ROC) curves for each identification results of articulation error patterns in training database were estimated. The equal error rates of Fronting, Backing, De-aspiration, Stopping, Affrication, and Omission were 7.32%, 11.78%, 9.87%, 8.76%, 7.07%, and 4.89%. Moreover, the thresholds with equal error rate were 0.16, 0.086, 0.092, 0.2, 0.2, and 0.1, respectively. For the testing database, the accuracy, specificity, sensitivity, and Kappa for each articulation error pattern were shown in Fig. 6.

3.2 Results of 3D facial animation

In this experiment, the spontaneous speech corpus with 3647 sentences was collected

to evaluate the performance of phoneme segmentation. Comparing proposed contextual knowledge based phoneme segmentation, Hotelling's T^2 test and BIC were applied as baseline systems. The experimental results of phoneme segmentation of those approaches were shown Fig. 7. It is clear that the contextual knowledge based phoneme segmentation with duration model can achieve best performance.

For 3D facial animation, the synthesized 3D tongue animation was also compared with 2D animation of mouth cavity and parts of animation were shown in Fig. 8. The experimental results achieve practical performance. Finally, the interface with 3D facial animation and text corpus for language training was shown in Fig. 9. In this approach, the text corpus for language training can be easily changed to be suitable for a user.

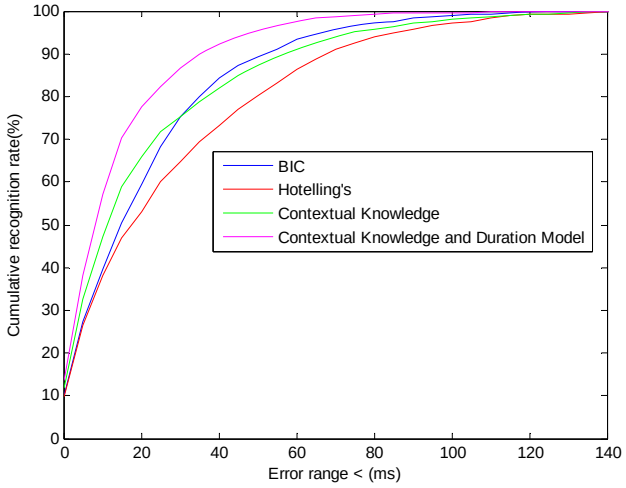

Fig. 7. Experimental results of phoneme segmentation

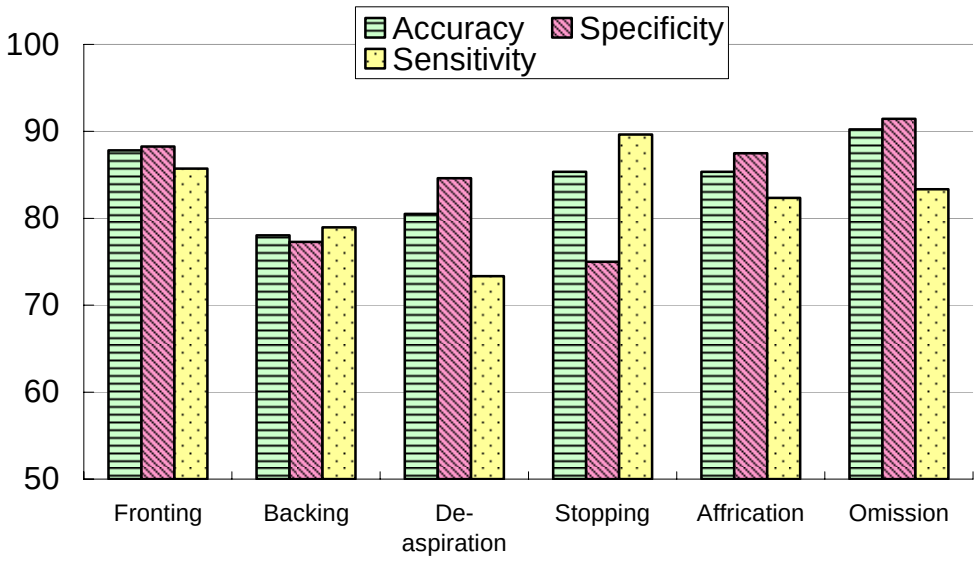

Fig. 6. The experimental results of identification of articulation error patterns

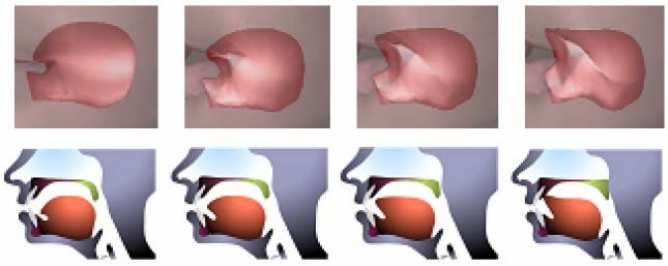

Fig. 8. Examples of synthesized 3D tongue animation and 2D animation of mouth cavity

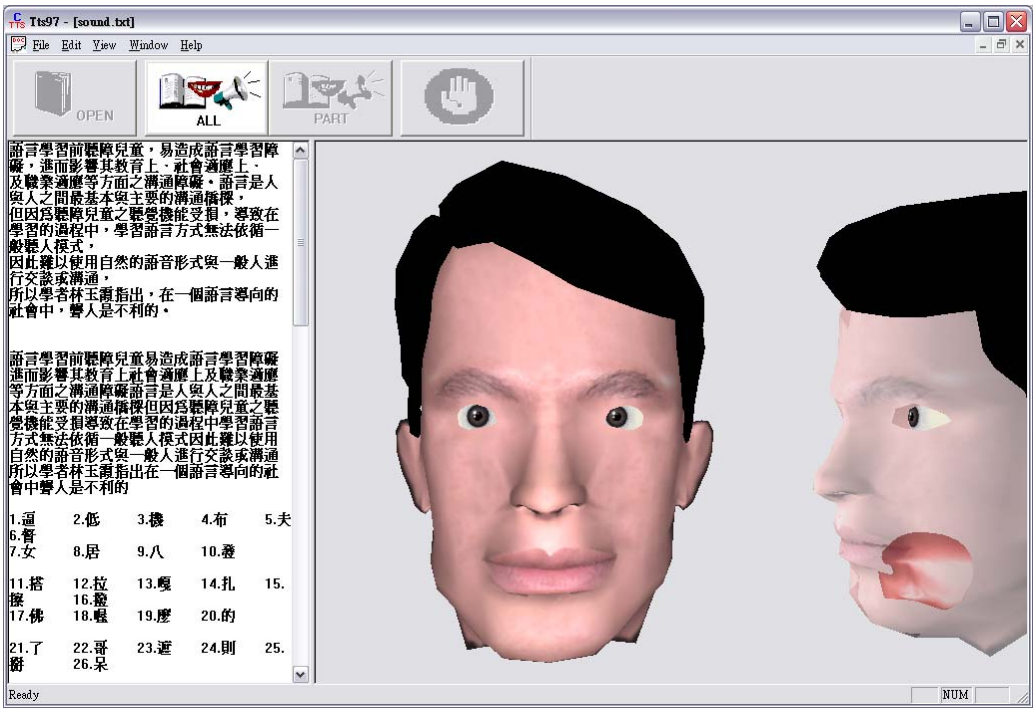

Fig. 9. The interface with 3D facial animation and text corpus for language training

4. Conclusion

This work had presented an innovative approach to identify articulation error patterns and 3D facial animation based language training. It can assist speech-language pathologists in clinical speech evaluation and training. Using spontaneous speech based interactive interface, the articulatory information of articulation disorders can be effectively and friendly acquired in speech signal. Integrating dependency network, the pronunciation errors and articulation error patterns can be automatically identified. Moreover, a 3D facial animation including speech signal, lip movement, and tongue movement can promote articulation disorder to simulate the articulatory behavior of tongue. The training corpus can be also easily modified to satisfy users articulatory ability. Besides, the language training activities in games is provided to caregivers to

enhance articulatory ability of articulation disorders in daily life. The experimental results show that this method is feasible.

ACKNOWLEDGEMENT

The authors would like to thank the National Science Council, R.O.C., for its financial support of this work, under Contract No. NSC 97-2221-E-218 -043.

REFERENCES

Ali, A.M.A., Van der Spiegel, J., and Mueller, P., "Acoustic-phonetic features for the automatic classification of stop consonants," *IEEE Trans. Speech and Audio Processing*, vol. 9, issue 8, pp. 833-841, Vol. 2001a.

Ali, A.M.A., Van der Spiegel, J., and Mueller, P., "Robust classification of stop consonants using auditory-based speech processing," in Proc. *IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 1, pp. 81-84, May 2001b.

Bernthal, J. E., and Bankson, W. B., Articulation and Phonological Disorders. Allyn and Bacon, 2004.

Georgoulas, G., Georgopoulos, V.C., and Stylios, C.D., "Speech Sound Classification and Detection of Articulation Disorders with Support Vector Machines and Wavelets," in Proc. *the 28th IEEE International Conference on EMBS Annual*, 2006.

Hueber, T., Aversano, G., Chollet, G., Denby, B., Dreyfus, G., Oussar, Y., Roussel, P., and Stone, M., "Eigentongue Feature Extraction for an Ultrasound-Based Silent Speech Interface," in Proc. *IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 1, pp. 1245-1248, April 2007.

Kramer, D. M., Hawryszko, C., Ortendahl, D.A., and Minaise, M., "Fluoroscopic MR imaging at 0.064 tesla," *IEEE Trans. on Medical Imaging*, vol. 10, issue 3, pp. 358-361, Sept., 1991.

Le Goff, B., *Synthèse à partir du texte de visages 3D parlant français*, PhD thesis, Grenoble, France, Oct. 1997.

Li, J., Tsao, Y., and Lee, C. H., "A Study on Knowledge Source Integration for Candidate Rescoring in Automatic Speech Recognition," in Proc. *IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 1, pp. 837-840, March 2005.

Mak, B., Siu, M.H., Ng, M., Tam, Y.C., Chan, Y.C., Chan, K.W., Leung, K.Y., Ho, S., Chong, F.H., Wong, J., and Lo, J., "PLASER: Pronunciation Learning via Automatic Speech Recognition", in Proc. of *HLT-NAACL 2003*, Edmonton, Canada, 23-29, 2003

Munhall, K. G., Vatikiotis-Bateson, E., and Tohkura, Y., "X-ray film database for speech research," *Journal of the Acoustical Society of America*, vol. 98, pp. 1222-1224, 1995.

Orzechowski, T., Izworski, A., Tadeusiewicz, R., Chmurzynska, K., Radkowski, P., and Gatkowska, I., "Processing of pathological changes in speech caused by dysarthria," in Proc. of *2005 International Symposium on Intelligent Signal Processing and Communication Systems*, pp. 49-52, Dec. 2005.

Preisach, C. and Schmidt-Thieme, L., "Relational Ensemble Classification," in Proc. *Sixth International Conference on Data Mining*, pp. 499-509, Dec. 2006.

Rajamanohar, M. and Fosler-Lussier, E., "An evaluation of hierarchical articulatory feature detectors," in Proc. of *2005 IEEE Workshop on Automatic Speech Recognition and Understanding*, pp. 59-64, Nov. 2005.

Robineau, F., Boy, F., Orliaguet, J.P., Demongeot, J., and Payan, Y., "Guiding the Surgical Gesture Using an Electro-Tactile Stimulus Array on the Tongue: A Feasibility Study," *IEEE Trans. Biomedical Engineering*, vol. 54, issue 4, pp. 711-717, 2007.

Sheng, H. et al, "Report of the police recommendation for manpower of linguistic therapist," *The Magazine of Hearing and Language, The Speech-Language-Hearing Association of the Republic of China*, vol. 16, pp 76-91, 2001.

Sher, Y.J., Chen, Y.J., Chiu, Y.H., Chung, K.C., and Wu, C.H., "MAP-based Perceptual Speech Modeling for Noisy Speech Recognition," *Journal of Information Science and Engineering*, vol. 22, no. 5, pp. 999-1013, Sep. 2006.

Siniscalchi, S. M., Schwarz, P., and Lee, C. H., "High-Accuracy Phone Recognition By Combining High-Performance Lattice Generation and Knowledge Based Rescoring," in Proc. of *IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 4, pp. 869-872, April 2007.

Stearns, M. and Frisch, S. A., "Production and perception of place of articulation errors," *Journal of Acoustic Society of America*, vol. 120, no. 5, pp. 3251, Nov. 2006.

Tian, Y., Yang, Q., Huang, T., Ling, C. X., and Gao, W., "Learning Contextual Dependency Network Models for Link-Based Classification," *IEEE Trans. Knowledge and Data Engineering*, vol. 18, issue 11, pp. 1482-1496, Nov. 2006.

Westbury, J. R., X-Ray Microbeam Speech Production Database User's Handbook, WI: University of Wisconsin Waisman Center, 1994.

Wu, S. L., Teaching activities of language disorder, National Kaohsiung Normal University Special Education Center, 2000.

Yu, B.L., "Appraisal and therapy of language disorder for children," *The Speech-Language-Hearing Association of The Republic of China*, pp.29-35, 1992.