ICICS-PCM 2003
15-18 December 2003
Singapore

Computer-based Malay Articulation Training for Malay Plosives at Isolated, Syllable and Word Level

H N Ting[1]; J Yunus[1]; S Vandort[2] and L C Wong[1]

[1]Faculty of Electrical Engineering, Universiti Teknologi Malaysia, Malaysia
[2]Department of Audiology and Speech Sciences, Universiti Kebangsaan Malaysia, Malaysia

Abstract

This paper describes the use of computer as an articulation training system for Malay plosives at isolated, syllable and word level. The proposed system is more convenient than the traditional speech analyzing tools such as Electropalatograph, where the latter requires an external electronic circuit to be attached into the mouth of client. The system is designed in a way that is user friendly and easy to use for the speech-language pathologist or even the client. The client undergoes speech training by just talking into the microphone and the system is able to recognize the sounds and classify them accordingly. Audio and visual feedback is used to help the client to identify his or her articulation errors as well as to make comparisons between his/her articulation models with the standard model. The system can be used for both children and adults.

1. Introduction

Articulation is basically speech sound production, which is the result from movements of the lips, tongue, velum and jaws to shape the flow of air into sounds [1]. The speech sounds are comprised of three groups: vowels, diphthongs and consonants. Vowels are produced with a relatively open oral cavity with phonation. Diphthongs are vowel-like sounds that are produced as a combination of two vowels. Consonant sounds are produced by constricting the airflow in the oral cavity or the nasal cavity.

Articulation problems are concerned with the failure in mastering the speech sound production in a spoken language. The causes of the articulation problems can be organic or functional. The organic causes include anatomical, motor or sensory impairments whereas the functional causes have various etiologies [2]. Consonant sounds are relatively more difficult to be produced than the vowel and diphthong sounds and hence most articulation problems involve consonant sounds. Thus, the research focuses on the articulation of consonant sounds instead of vowel sounds.

In conventional speech training, all kinds of material can be used as tools to help treat the articulation problem. These materials include a candle (to detect bilabial sounds), peanut butter (to detect alveolar and velar sounds), a mirror (to provide visual feedback), a tongue depressor (to help cue the client on correct positioning of the tongue), cosmetic mirrors (detect breath vapour), cotton ball (to detect lingua-dental sounds) and so on. The use of these devices is welcomed because they are cheap in cost and the client can produce the speech sounds in a natural manner without any devices connected to his/her mouth. However, this kind of training may require total involvement of a speech-language pathologist (SLP) to monitor and evaluate the client's speech production performance, where the job is quite laborious and repetitious.

Special speech devices such as the Electropalatograph (EPG) and spectrograph are now used in analyzing articulation problems. EPG is simply an artificial palate put into the mouth of the client to detect the movement of tongue during speech production [3]. The major drawback of this device is that the client is not able to produce the speech sounds in a natural manner during the speech training session. The spectrograph is used by the SLP to analyze the speech production pattern of client. However, its time-frequency analysis is not user-friendly to the client because the analyzing result is difficult to be interpreted by the client.

Realizing the problems and limitations of the conventional articulation training and speech analyzing tools, thus we propose a computer-based articulation speech training system to be more user-friendly and easy for use. The proposed system eliminates the use of candles, feathers, cotton balls and mirrors in the training sessions. It allows the client undergoes the articulation training by just talking to the microphone, which is connected to the computer system. The system provides the audio and visual feedback to help the client to identify his articulation pattern. The system enables the training session to be conducted without the presence of SLP, thus erasing the SLP's laborious and repetitious job of monitoring.

The use of computer technology in speech therapy and assessment is still quite new in Malaysia. The current available computer-based speech therapy systems in the Clinic of Audiology and Speech Sciences, Universiti Kebangsaan Malaysia (UKM), Malaysia are the IBM Speech Viewer [4] and Kay Elemtrics VisiPitch. These systems are designed mainly for voice therapy and are not

for the purpose of articulation therapy or training. Furthermore these systems are English-language-based. The motivation of the research is to develop a Malay-language-based articulation training system to be used in Malaysian environment. The proposed system is able to conduct the articulation for Malay plosives at isolated sound, syllable and even at word level. In Malay language, the plosives are comprised of [b], [d], [g], [p], [t] and [k].

Section 2 describes the graphical user interface (GUI) of the proposed system. Section 3 shows and discusses the system tests and results of the proposed system, which involves adults as well as children speakers. The paper is concluded in Section 4. Acknowledgements and references are shown in Section 5 and Section 6 respectively.

2. System Description

The computer-based Malay articulation training system consists of training at three levels of Malay plosive sounds: isolated sound, syllable and word level. The GUI of each level is shown in Figures 1, 2 and 3 respectively. All levels of the training basically share the similar GUI design except the lists of target sounds and their related pictures.

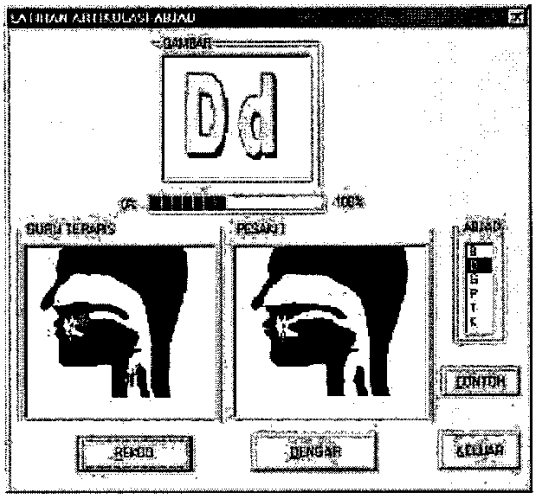

Figure 1: Articulation training at isolated sound level

The user/client undergoes the articulation training simply by just talking into the microphone, which is connected to the computer system. First, the client selects the target sounds from the list. The selection can be done sequentially or in a random manner. The picture of the target sounds will be displayed according in the group box of "Gambar". The standard model of the target sound is shown in the group box of "GURU TERAPIS". The client clicks on the "REKOD" button to start speaking. The progress of the recording is shown by the progress bar, which indicates the start and end of recording. The visual feedback of the client then will be displayed in the group box of "PESAKIT". The client could playback his/her recorded voices by simply pressing the "DENGAR" button. The standard sounds can be heard by pressing the "CONTOH" button.

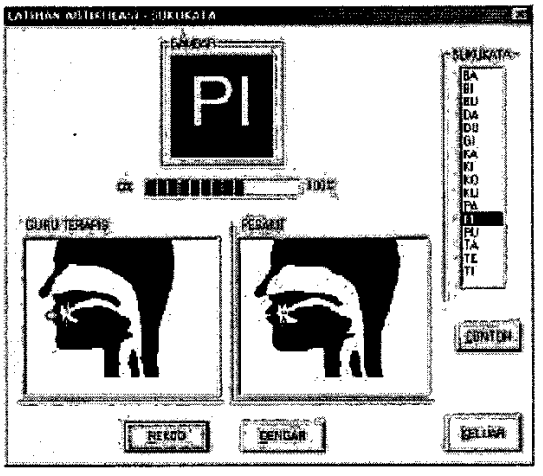

Figure 2: Articulation training at syllable level

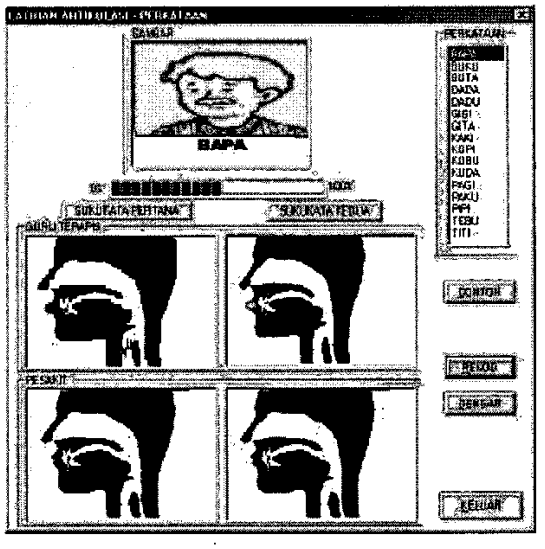

Figure 3: Articulation training at word level

The models of the speech production are displayed in the pictures of the saggital view of a human vocal tract. A colored circle is used to indicate the place of articulation of the target sounds. The voicing part of the target sounds is differentiated by the circle colour and waveform signal at the glottis. Voiced plosive sounds are featured by red-colour circle with the presence of red-colour waveform

signal. The green-colour circle with absence of waveform signal indicates the unvoiced counterparts.

The list of isolated plosive sounds includes [b], [d], [g], [p], [t] and [k]. The list of plosive syllable sounds contains 16 syllables: [ba], [bi], [bu], [da], [du], [gi], [ka], [ki], [ko], [ku], [pa], [pi], [pu], [ta], [te] and [ti]. At word level, the 16 Malay words are "*Bapa*" (father), "*Buku*" (book), "*Buta*" (blind), "*Dada*" (chest), "*Dadu*" (dice), "*Gigi*" (teeth), "*Gita*" (song), "*Kaki*" (legs), "*Kopi*" (coffee), "*Kubu*" (fort), "*Kuda*" (horse), "*Pagi*" (morning), "*Paku*" (nail), "*Pipi*" (cheek), "*Tebu*" (sugar cane) and "*Titi*" (small bridge). These words allow the consonant sounds to be produced at the initial and medial position of the words.

3. System Tests and Discussions

Several system tests have been conducted on children and adults. The children were comprised of students in a primary level religious school at Universiti Teknologi Malaysia (UTM). The involved adult speakers were from the UTM's undergraduate students.

3.1 System Tests on Children

The system tests included the tests at isolated sound, syllable and word levels. Two different groups of children were involved in the system testing. One group was involved at the isolated level test and the other was involved at both syllable and word level test.

3.1.1 Isolated Sound Level

The test involved five normal children with their age between nine and ten years old. Each child was asked to speak into the microphone five times for each plosive sound. The result was shown in Table 1.

The overall accuracy of the system was about 59%. The system was excellent in recognizing [b], [p] and [k] sounds. However, the system failed to recognize the [d], [g] and [t] sounds. This might be due to inaccurate edge detection of the speech sounds by the computer. The system's performance can be improved by increasing the amount of speech samples used to set up the speech recognition system and by increasing the accuracy of the system in detecting the edge points of the speech sounds

Table 1: Test result at isolated sound level on children

Sound	Recognized as							Acc.
	B	D	G	P	T	K	Others	(%)
B	18	2	0	0	3	0	2	72.0
D	3	9	1	1	5	0	6	36.0
G	1	0	9	1	1	4	9	36.0
P	0	1	0	0	24	0	0	96.0
T	0	0	0	6	12	6	1	24.0
K	1	0	0	23	0	1	0	92.0

3.1.1 Syllable Level

The test subjects were ten children aged between seven and ten. The group consisted of six males and four females. Each child was told to utter each syllable sound once, which resulted in a total number of 160 sample sounds.

The system was able to achieve an overall accuracy of about 88%, which was 28% better than the performance at isolated sound level. The detail information about can be found in Table 2. Obviously, the system was able to recognize the speech sounds in a speaker-independent manner.

3.1.1 Word Level

Each of the ten children was requested to utter each word once. Each of the word was comprised of two plosive syllable sounds. Thus, the total number of target speech sounds was 320. The test result was shown in Table 3. The overall performance of the system was about 76%. From the test results, it appears that the system is more suitable to be used for articulation training at syllable level rather than at isolated sound and word level.

Table 2: Test result at syllable level on children

Sound	Recognized as							Acc.
	B	D	G	P	T	K	Others	(%)
B	29	1	0	0	0	0	0	96.7
D	0	18	0	1	0	0	1	90.0
G	0	0	8	0	0	0	2	80.0
P	0	0	1	26	3	0	0	86.7
T	0	3	0	0	24	1	2	80.0
K	0	0	1	1	0	36	2	90.0

Table 3: Test result at word level on children

Sound	Recognized as							Acc.
	B	D	G	P	T	K	Others	(%)
B	37	7	0	4	1	1	0	74.0
D	4	44	0	0	0	0	2	88.0
G	2	0	33	0	0	2	3	82.5
P	5	2	1	37	7	2	6	61.7
T	4	3	2	1	36	1	3	72.0
K	1	2	2	5	2	55	3	78.6

3.2 System Tests on Adults

Only one group of five adult speakers were involved in the system tests at both syllable and word levels. The tests were conducted in normal room environment.

3.2.1 Syllable Level

Each speaker uttered each of the syllable sound once, resulting in a total number of 80 speech samples. The test result was shown in Table 4. The overall accuracy of the system was about 89%, which was comparable to the children's test result. It again suggested that the system was able to recognize plosive syllable sounds of both children and adult in speaker-independent manner.

Table 4: Test result at syllable level on adults

Sound	Recognized as							Acc.
	B	D	G	P	T	K	Others	(%)
B	15	0	0	0	0	0	0	100.0
D	1	8	0	0	1	0	0	80.0
G	1	0	3	0	0	1	0	60.0
P	0	0	0	14	0	0	1	93.3
T	0	0	0	3	11	0	1	73.3
K	0	0	0	0	0	20	0	100.0

3.2.1 Word Level

Each speaker uttered each word once and the total number of speech samples was 160. The test result was shown in Table 5. The overall performance of the system was 80%. The test result suggested that the system was more appropriate to be used at syllable level than the word level.

Table 5: Test result at word level on adults

Sound	Recognized as							Acc.
	B	D	G	P	T	K	Others	(%)
B	24	1	0	0	0	0	0	96.0
D	6	16	0	0	2	0	1	64.0
G	2	0	15	1	0	1	1	75.0
P	1	0	1	27	0	1	0	90.0
T	1	2	2	3	16	0	1	64.0
K	0	1	1	1	2	30	0	85.7

4. Conclusion

A computer-based Malay articulation training system has been developed for Malay plosive sounds at isolated sound, syllable and word levels. The system can be used as the articulation training system for both children and adults. The training session becomes more natural where the client undergoes the training simply by just talking to a microphone. No external circuits are attached into the client's mouth. The system provides audio as well as visual feedbacks to help the client to achieve his/her target sounds. The system is user-friendly and is easily used by both the SLP as well as client.

References

[1] P.J. Thomas and F. F. Carmack, Speech and Language Detecting and Correcting Special Needs. MA: Allyn & Bacon, 1990.

[2] C. Van Riper and R. L. Erickson, Speech Correction: An Introduction to Speech Pathology and Audiology. 9th ed. MA: Allyn & Bacon, 1996.

[3] A. A. Wrench, A. D. McIntosh and W. J. Hardcastle, "Optopalatograph (OPG): A New Apparaturs for Speech Production Analysis," Proceedings of the Fourth International Conference on Spoken Language (ICSLP96), Vol. 3, pp. 1589 - 1592, 1996.

[4] F.R. Adams, H. Crepy, D. Jameson and J. Thatcher, "IBM Products for Persons with Disabilities," Proceedings of the Global Telecommunications Conference and Exhibition (GLOBECOM'89), Vol. 2, pp. 980 - 982, 1989.