MENU
PRINT VERSION
HELP & FAQs

Visual Teaching Aid for Pronunciation Acquisition

Yuko Hoshino
Norio Sato
Kazumasa Kawai

Kanazawa Institute of Technology

We have developed three-dimensional computer graphics to show learners the movement of the mouth when doing Mandarin Chinese pronunciation. Unlike videotape, computer graphics can be translucent, and allows learners to see inside the mouth through the skin. In addition, graphics can allow deformation, so that the critical features may be emphasized. The learners are provided audio guidance, in Japanese, of how to produce each Chinese sound element. In addition, they can see the interactions of the lips, teeth, and tongue from different angles as they listen to the sound, and can mimic the movements at their own learning pace. This process is repeatable. We then examined its effectiveness through observing the learning process and evaluating the learners' pronunciation changes. The tool has proven to be particularly effective for false beginners to review and brush up their pronunciation as well as for true beginners.

本論文は3次元コンピュータグラフィクス(CG)を使用した中国語(普通話)発音学習支援教材の開発と、その教育への応用について論ずる。ビデオ教材と異なり、CGは半透明にすることができ、学習者が中国語の発音時の口腔内の唇や舌の動きと歯の位置との関連を目で見ることができる。さらにデフォルメによって発音に重要な動きを強調して見せることができる。本教材では、学習者は任意の方向からのイメージを見ながら発音を聞いて口の動きをまね、自分の学習ペースで発音学習ができる。学習は何度も繰り返し行える。教材の有効性について学習者の学習プロセスの観察と発音の変化により評価を行ったところ、初学者の発音習得に有効なだけでなく、過去に学習したことはあるが発音を忘れてしまった再学習者の学習支援に特に有効であることがわかった。

Introduction

When learning a new language, acquiring the skills to produce new and unfamiliar sounds poses a significant challenge. All language educators will agree that this is a time-consuming process. What if we could show learners the way native speakers produce the sounds of their languages? What if they could learn the pronunciation at any time and at any place?

There are many video materials available to learn Chinese pronunciation. One drawback of such video materials is that one cannot see the movement inside the mouth, i.e. the movement of tongue, interaction of tongue and teeth and so on. Also, it is difficult to pinpoint the part of the video to see the movement one wants to learn.

To overcome this shortcoming, we have developed computer graphics to show learners of Chinese the movement of Chinese pronunciation by three dimensional computer graphics. We created graphic elements by using modeling software and made them controllable by programming. The learners are provided with the audio guidance of how to produce each Chinese sound element, and can see the relationships among the movements of the lips, teeth, and tongue as they listen to the sound. They can see the parts of the mouth from different angles by handling the mouse and can mimic the movements at their own learning pace. This process is repeatable at learners' convenience.

Sound systems

English and Chinese sound systems

Japanese students learn English as a foreign language at Junior-high and high schools. Thus, most of the students have acquired some degree of English pronunciation ability. It has been observed that some of them transfer their knowledge of producing English sounds to produce Chinese sounds, which is sometimes helpful. One such example is the "er" of Chinese and English "r." However, due to other unfamiliar sound elements in Chinese such as "e" and "qu," many Japanese speakers find it difficult to learn the new phonological system of Chinese.

Japanese and Chinese sound systems

PingYin transcription

Although Japanese speakers and Chinese speakers share Chinese characters in their writing system, the Chinese pronunciation system poses many challenges for Japanese speakers. First, the PingYin transcription using alphabets to teach Chinese sounds can be difficult to read intuitively. Typical examples are "xi" and "qu" which Japanese learners would not see in their native language.

Multiple vowels

Because Japanese sounds mostly consist of one consonant and one vowel, Japanese learners of Chinese tend to reduce multiple vowels of Chinese into a single vowel. Therefore, it is critical for them to practice correctly and repeatedly until producing the sounds become automatic. With limited classroom instruction time and availability of instructors, computer software is ideal for learners to practice at their convenience and with their own learning speed.

Table 1. Japanese syllables and Chinese syllables

Type of syllables	Japanese	Chinese (ping-yin)
vowel	5 (a, i, u, e, o)	7 (a, i, u, ü, e, o*, er)
Consonant + 1 vowel	ka, hi, mo, etc.	ka, xi, qu, etc.
Consonant + y + vowel	kya, hyu, myo, etc.	-
Consonant + 2 vowels	-	kai, xia, qie, etc.
Consonant + 3 vowels	-	kuai, xiao, qiao, etc.

* Same alphabet letter does not necessarily represent the same pronunciation.

The same alphabet in Table 1 does not necessarily represent the same sound between Japanese and Chinese. In addition, due to its familiarity, Japanese speakers often use Katakana transcription instead of the PingYin in an attempt to capture the Chinese sound. Nevertheless, “shi” sound and “xi” sound would be transcribed as “シ”, “zhi” sound and “ji” sound as “ジ”, “chi” and “qi” as “チ.” It is almost impossible to differentiate those sounds if using Katakana transcription. As Yan and Liu (1997) suggest, learners should associate the PingYin with Chinese sounds, avoiding associating with alphabetical transcriptions of other languages.

Aspirated and un-aspirated sounds

Distinction in aspirated and un-aspirated sounds in Chinese is very important in learning Chinese pronunciation. This is also problematic since Japanese speakers do not make such distinction as clearly as Chinese speakers do. Many Japanese speakers have trouble hearing the difference, thus have also trouble differentiating these two types of sounds. In this suit, we indicate aspiration as a white arrow.

Tones

Another important aspect of Chinese pronunciation is tones. Tones are produced by vibration of vocal cord and not by the shape of mouth. Showing tones should be left for other means of representation rather than computer graphics.

Learning tool development

Requirements

From the functional, characteristic and technical point of view, our design criterions are as followings:

(1) We have decided to create a woman to show the movement of mouth. The reason is that it could be easier for learners to make rapport with a friendly-looking woman, as depicted in Figure 1. In addition, women’s speech is usually clearer than men’s speech.

Figure 1. Model image.

(2) We should take full advantages of three dimensional computer graphics features that other medias such as video graphics do not possess.

(3) The tool should be widely available in school campus. Our target is the Note PCs that teachers and students have individually. Only feasible and available technologies should be used in practical application.

Fundamental features

Compared with videotapes, three-dimensional computer graphics have three fundamental differences: it allows for translucency, view from different angles, and deformation.

Translucency

Computer graphics can be translucent. As shown in Figure 2, our suite allows learners to see inside the mouth through the skin. Our tool is not, however, for medical use, and so the translucency need not show potentially shocking anatomical details. One remedy is semi-translucency, not pure transparency. From some angles, when this feature is not useful, even the translucency may be avoided. This issue is presently under further study.

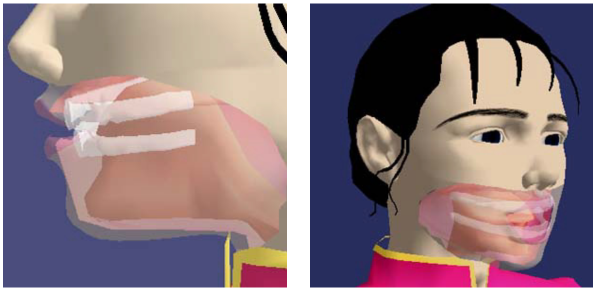


Figure 2. Translucent views.

Arbitrary Angles

Computer graphics allows for viewing from arbitrary angles on one model. Users can choose any angle by turning the model with a mouse or keys. Repetition of the same angle shows only a part of shapes, and may become boring. As shown in Figure 3, we think that at a minimum, side and diagonal views as well as front view are useful.

Side angle **Upper diagonal view**

Figure 3. Views from arbitrary angles

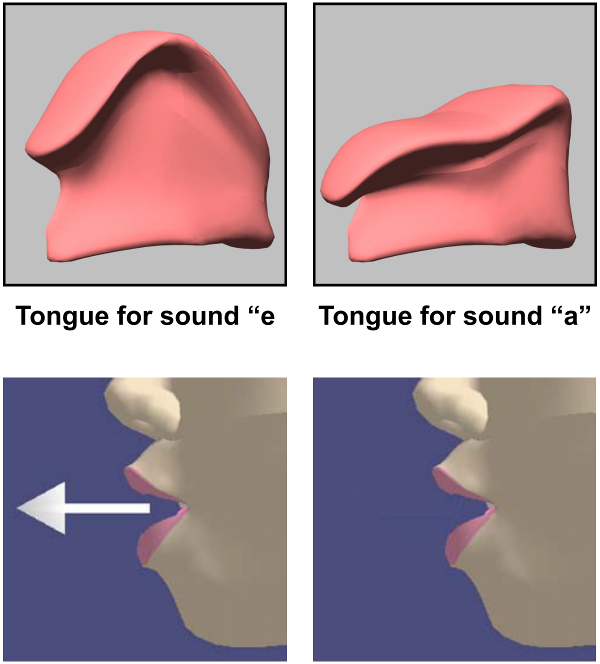
Deformation and emphasis

Critical features may be emphasized. As shown in Figure 4, sometimes, it is useful to show the movement of lips and the tongue in exaggerated forms rather than real forms. We checked such forms for each sound element considering the differences from or missing in Japanese sound elements. We also attached an arrow to visualize breath, when emphasis is needed.

Implementation

Step 1: Capturing human video and voice

One of the authors, who is teaching Chinese to undergraduate classes, was the model. It is an advantage in that she has experiences and knows crucial points of each sound for Japanese (or other foreign) learners. Using a conventional video camera and a PC, we captured for each sound element her voice and shapes from two or three angles.



Tongue for sound "e

Tongue for sound "a"

Aspirated sound "po"

Un-aspirated sound "bo"

Figure 4. Deformations and emphasis.

Step 2: Creating the CG model

We used a polygon based modeling software LW (LightWave3D®) to create the character. LW is suitable for organic objects such as human or animals rather than precise modeling such as architectural CAD. Thanks to its features, e.g., "sub-patch" and "V-map", LW enables us to save the number of polygons to generate smooth surfaces. This facilitates the transformation of models.

We created the head by putting together eyes, ears, the nose, and the mouth to the face surface. We then divided the face surface into two parts: one is the cheek and the other is the rest parts of the face, so that we may show the inside of the mouth visible by changing the cheeks to semi-translucent.

The mouth consists of movable parts: tongue, teeth, and the roof. These parts are independent so as to be separately moved by programming.

Step 3: Adding motions to the model

Morphing technique: Adding motions for the sound element can be realized by "morphing" technique. Morphing is a programming task. Morphing requires two or more key frames, while programming enables automatic generation of complementary frames between key frames. An example is shown in Figure 5.

Producing key-frames: Producing a key frame is rigorous work. Observing the video image, the movements of the jaw, the lips tongue and teeth were shaped. The movements of the lip and the tongue were exaggerated a bit, when it was thought desirable to do so. First, outlining the transformation of the face is done by rotating the jaw, followed by the transformation of the muscles, Iijima (2000), which determine facial expressions. Inside, the mouth is shaped according to vertical sections published by Chinese Language Friendship Society and Kuraishi (1984).

For each simple vowel, the number of key frames is two, i.e., mouth shut to the vowel specific frame. For each double vowel, three key frames are required, and for each triple vowel four, etc. We need,

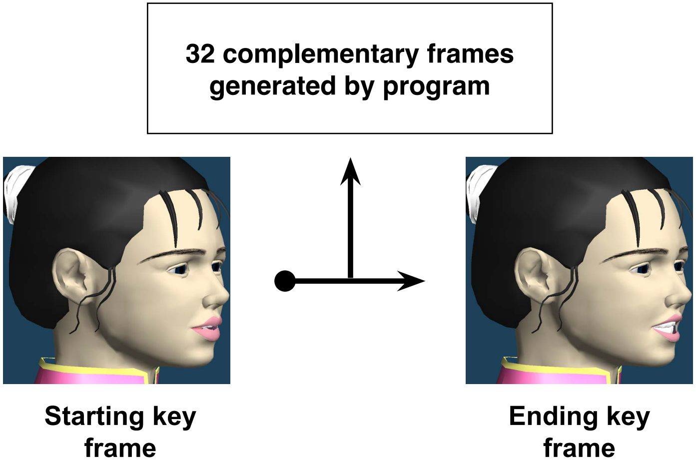


Figure 5. Animation by morphing technique.

therefore, seven key frames for the simple vowels and twenty-one key frames for the consonants, etc. We can save six consonants, since a pair of an aspirated and an un-aspirated consonants share the same shape. Some signs can indicate an aspiration by program. We combine simple two or more vowels for composite vowels, and for syllables, we combine consonants with these vowels.

Step 4. Controlling animation and voices

To control the animation and voices on Windows PC, we used a conventional programming tool set, i.e., Direct X® library and Visual C++®.We adjusted the morphing speed to the voice speed, because the former is dependent on PC performance while the latter is almost independent. LW produces key frames as "X-file" suited for Direct X® library. This file contains one or more objects each of which consists of many object vertices and mapping of the texture attached.

Figure 6 shows a snapshot of the developed tool:

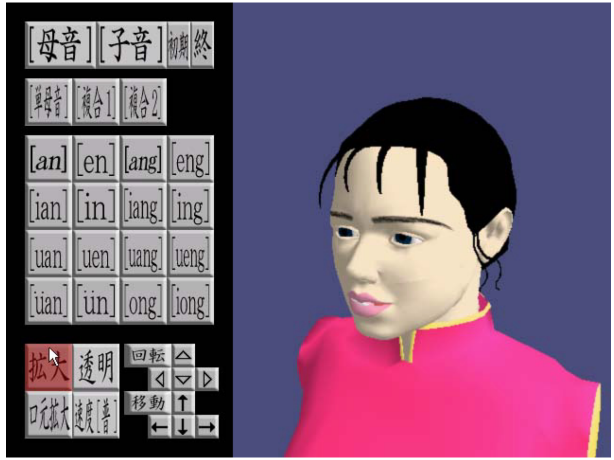


Figure 6. UI (User interface) snapshot

(1) User Interface: The left hand side is the control panel, which consists of menu, sound, and CG control buttons.

(2) Selection of sound sets*:* Pushing a menu button to specify a set of sound elements starts the loading of a set of key frames. Then, the buttons for each sound element in this set appear.

(3) Playing & Learning: Adjusting the mouse cursor to a sound button, and click the left button of mouse, the animation and pronunciation voice starts, while the right button of mouse produces explanation voice. The CG control button allows for changing the position, angles and translucency as one wish.

Application

Target users and curricula

Learners that we assume this tool will be useful to would be in three categories:

A. "True" beginners (learners without any previous knowledge or experience)

B. "False" beginners (learners with some previous knowledge and experience)

C. Learners of self-study, such as students at the University of the Air (Hoosoo Daigaku) and learners using TV or radio language programs

Learning process

Typical steps for learners in any Chinese course in learning pronunciation are as follows.
Step 1: learning single vowels*
Step 2: learning double- and triple-vowels*
Step 3: learning consonants*
Step 4: combining consonants and vowels
*Learning vowels and learning consonants are interchangeable.

The tool follows these steps, but learners can start the program from any step depending on their previous knowledge, experience, and needs in learning Chinese.

How to use the tool

Learners can use the tool in two ways. (1) It can be used for self-study and (2) can also be used in class with teacher instruction. The interface of the tool is quite self-explanatory. It has a button for each sound element and if a user left-clicks, a sample sound is played and if right-clicks, the instruction of how to produce the sound is played. The user can proceed to hear every sound and can repeat the sound as many times as s/he desires. Although it is advisable that the use repeats the sound to acquire correct pronunciation, repeating may not be necessary if the purpose of study is for listening comprehension.

Evaluation

We have asked volunteers of true beginners, false beginners, and self-study learners to use this tool for a week and evaluated their changes in pronunciation. As for true beginners, if used as a supplementary reviewing tool besides classroom instructions, this tool is helpful in reducing necessary class time for practice. False beginners had taken Chinese lessons but had forgotten how to pronounce PingYin at the time of the evaluation. "e" "ü" "er" "üe" "ün" "eng" "uen" are among the most difficult sounds for them to correctly reproduce. This is probably due to the discrepancy between intuitive reading of PingYin and actual sounds. However, when these learners used the tool to listen to the sounds and explanations while watching the graphics, they quickly reproduced the correct sounds and maintained them. In the interviews after the trial, they commented that the combination of graphics and explanations helped them to produce the sounds. The self-study learners commented that because of the tool's interactivity, learning was enjoyable.

The tool can be used not only at the beginning of any Chinese course but also at the beginning of each term of Chinese program as review and reinforcement.

This tool may also be applied to learning pronunciation of other languages, such as English and German.

Conclusion

Pronunciation accounts for a considerable part of Chinese language study due to its complicated sound system. To overcome the shortcoming of conventional learning aids such as videos and CDs, we have developed a three-dimensional computer graphics that enables the views of translucency, different angles, and deformations of parts of the mouth.

We named the suit "YinJie" meaning "Sound Lady" in Chinese. The suite runs on Windows PC without additional software. We also consider installing this suit for e-learning. Due to the limitation of present technologies, more widespread use such as with mobile phones is left for future work.

References

Yan, M.M. and Liu, J. L. (1997). *Interactions I*. Indiana University Press

Iijima, T. (2000). *Anatomy of the Human Body* (Jintai no shikumi). Work Corporation

Chinese Language Friendship Society and Kuraishi, T (1984). *Chinese Pronunciation Class*. Revised edition, Taishukan Publishing