ðə dʒərnəl əv læŋgwədʒ titʃɪŋ ænd lərnɪŋ

J L T L

dɛdəketəd tu haj kwələti risərtʃ ɪn əplayd lɪŋgwɪstɪks

http://www.jltl.org/jltl/

The Journal of Language Teaching and Learning, 2015–1, 52-67

Using voice recognition software in learning of Chinese as a foreign language pronunciation

Herbert Mushangwe[1]

Abstract

This paper is based on the notion that what matters most in language learning is communication. We argue that if a learner can speak what can be understood by native speakers while at the same time understanding what the native speaker says, then we can conclude that this learner has successfully acquired the target language. We present a set of ideas based on the voice recognition exercise which was done with 20 students from different African countries who had studied Chinese as a foreign language for three months at Hebei University. A voice recognition application on smart phones and tablets was used for pronunciation practice purpose in order to boast students' confidence in spoken Chinese. In this exercise the phone acts as the decoder and it represents how a native Chinese speaker would decode the learner's speech. The results showed that during the exercise students become more and more conscious of their errors and adjusted their pronunciation. In addition, the voice recognition application helped to determine whether or not the learner's pronunciation was stable. Based on the results of the study, we concluded that it was possible to turn voice recognition application on smart phones into a game that language learners could use to practice their spoken Chinese. This method is cheap and it promotes self-evaluation as well as boosting interest in learning Chinese language.

Keywords: Voice recognition, Chinese as a foreign language, Pronunciation practice, Confidence, Stability.

1. Introduction

Voice recognition which is also well known as speech recognition or automatic speech recognition is now a common application on modern smart phones and tablets which among other functions can translate human speech into written words. Google's Android and Apple's iOS are among the most popular voice recognition applications. Voice recognition is defined as a process by which human sounds, words or phrases are converted into electrical signals which are then transformed into coding patterns to which meaning has been assigned such that spoken word can be used as an input to a computer program (Rabiner & Juang, 1993). In recent years this computer program has become available on cell phones, thus allowing people to use voice input to send messages on their phones even when they are driving. The primary purpose of voice recognition application on cell phones is for safety when sending messages while driving. However, voice

[1] Hebei University, 071002, Baoding Email: htsungy@yahoo.co.uk

recognition also makes use of cell phones more entertaining especially when it is used in such games as the TalkingTom application where one speaks to the phone and the rabbit (Mouse) repeats the speakers' words. Voice recognition applications have so many potential functions other than safety and entertainment some of which have not yet been explored. Gales and Young (2007) summarized the possible functions of voice recognition application as follows; commanding and controlling, dictation, transcription of recorded speech, searching audio documents and interactive spoken dialogues. All these functions of voice recognition can be also effectively implemented as teaching aids when teaching languages. Application of speech recognition in language learning is not anything new; Rolandi (2005) once said that Speech recognition can provide the means to interactively evaluate the utterances of a learner on several educational dimensions.

In this paper we attempt to demonstrate how voice recognition can be integrated into learning Chinese as a foreign language. Learning Chinese as a foreign language is becoming popular due to China's rise as one of the world's economic powerhouses (Abeysinghe & Lu, 2003). However, though many people are now studying Chinese, this language is considered as one of the most difficult languages in the world (Rosenberg, 1979; Shen, 1958). Considering the fact that it has so many dialects, and that Chinese language learners are people from different countries with different first languages, to achieve a unified pronunciation among these language learners is something that is quite challenging, if not impossible. Thus, in this paper we argue that what matters most in teaching Chinese as a foreign language is for the learners to be at least comprehended by Chinese native speakers rather than trying to make each and every student produce Chinese sounds like the teacher or the native speakers. This is not to say that listening ability is not important, neither is this a way of promoting non-standard Chinese pronunciation. In this paper we argue that language learners can use voice recognition software to determine if their pronunciation can be heard by others. Once the voice recognition device is able to transcribe the correct pronunciation of the learner, then it means that the learner's pronunciation can be understood by others.

2. Literature Review and Brief History of Voice recognition

Voice recognition technology is not a recent development; it has a long history that can be traced back into the remote past and up to the present a lot has been written about its application to language learning. According to Hsin and Hedge (1999) speech recognition programs came into market at the end of 1997, as a way of translating spoken speech directly into text. However, Juang and Rabiner (2004) tracks the origin of voice recognition to the invention of recording devices by Alexander Graham Bell and his cousins Chichester Bell and Charles Sumner Tainter in 1888. The history of voice recognition experiments documented by Juang & Rabiner shows that the success of modern day voice recognition is a product of researches over years. Anusuya and Katti (2009) notes that voice recognition has over 60 years of research but still needs some developments especially in terms of defining various types of speech classes, speech representation, feature extraction techniques, speech classifiers, database and performance evaluation.

Though voice recognition application still need further improvements, but still it has an important role in various fields in language learning. Furui (1999) classifies the role of voice recognition into dictation and human-computer dialogue systems. Ehsani and Knodt (1998) argue that; although use of speech technology in computer-based systems has its own limitations, however it stimulates interactive learning. Application of speech recognition technology in language learning started in the late 1990s. Zechner, Bejar and Hemat (2007) describe how they used speech recognition device in an experiment to assess pronunciation of non native speakers of English. Their experiment showed that speech recognition devise could reliably capture some aspects of speaking proficiency. In their acknowledgements the above researchers acknowledge the contribution of Multimodal Technologies Company in Pittsburgh and in particular Thomas Polzin (computer specialist) who helped them in explaining the use of speech recognition technology and helping them with data

transcription. This shows that the past voice recognition technology required highly qualified computer expertise.

Apart from that, some language learning softwares, such as Babbel language learning software which is based on smart learning system also introduced voice recognition software for pronunciation practice in 2009. This kind of speech pronunciation practice allows learners to have "instant evaluation of how close their pronunciation is to that of a native speaker" (O'Hear, 2010). When the learner speaks to the computer s/he is rated ranging from 0-100, with a 50 or higher meaning that s/he is understandable by the native speakers of the target language. According to O'Hear (2010), this kind of practice helps students of foreign languages to have more time to practice their speaking and at the same time it boasts confidence.

All these programmes are available on computers and such softwares are not available for free, this makes it difficult for language learners from developing countries to use them. Sangani (2013) notes that though fully developed speech recognition software was invented in 1975 but it was only available on computers for all these years. He further notes that use of voice recognition application on cell phones is a recent development, for instance Google Voice Search became available on Apple mobile devices only in October 2012. The past speech recognition devices were not only expensive, but also not locally available especially in developing countries. Following the fast technological advancement, voice recognition is now a common application on cell phones and anyone can be able to use them even without deep understanding of the technical know-how as it used to be in the past. In this paper we will demonstrate how the voice recognition software on cell phones can be integrated into learning of Chinese as a foreign language. The main aim of this paper is to help Chinese as a foreign language learners to be able to utilize this application which is now available on almost all smart phones yet the majority of people use it for entertainment only.

This research is based on the assumption that first language influence in second language learning is unavoidable. We base our argument on the notion that no learners can reach ultimate native-like pronunciation in second language (Hyltenstam & Abrahamsson, 2000; Scovel, 1988). Although we agree that first language influence hinders second language learners from achieving an ultimate native like pronunciation (Bada, 2001; Cook, 1991; Cummins, 1984) we also believe that the learner can still achieve proficiency level where communication between a native speaker and the learner is possible without any difficulties. Counselman (2010) argues that even though native-like pronunciation is difficult to achieve, learners can still approximate pronunciation which is as close to the native speaker as possible. Levis John and Kimberly LeVelle (2010) also observe that the main aim in teaching pronunciation is not to achieve a native-like pronunciation but to help learners achieve a comfortably intelligible pronunciation.

Learning second language requires constant practice which helps to build learner's confidence. In the classroom pronunciation practice the teacher is the key person who guides students' pronunciation. However, Laroy (1995) notes that correcting learners' pronunciation is one of the most frustrating aspects when teaching a foreign language, and at the same time it can be counter-productive since continuous correction of learners' pronunciation may be a major cause of loss of learners' confidence. In light of the above, when learners practice pronunciation it is common that sometimes the listener (the tutor) might become frustrated and fail to correct the learner's pronunciation errors, and at the same time the learner might get discouraged by constant corrections thus leading to loss of confidence. Therefore, we suggest the use of voice recognition software for pronunciation practice which involves a cell phone in place of human listener. The cell phone in this case substitutes the human listener and it acts as the evaluator of the student's pronunciation. The cell phone does not correct the student's pronunciation; rather it transcribes what the student says.

3. Standard Chinese pronunciation

Standard Chinese (also called Mandarin Chinese) is based on the pronunciation of the Beijing dialect. According to Duanmu (2007), Standard Chinese is spoken by over one billion people, but less than 1% of them do so without some accent and even the Beijing natives speak Standard Chinese with an accent. Considering the fact that speaking pure Standard Chinese is quite difficult even for the native speakers of Chinese, the question arises or whether or not it should be obligatory for students studying Chinese as a foreign language to produce Chinese words or phrases exactly like the Chinese people themselves. In teaching Chinese as a foreign language, Standard Chinese pronunciation is used as the bases for teaching despite the reality that 99% if not 100% of the learners will end up speaking Chinese with an accent.

This does not mean that teaching of Chinese is not successful rather it is a sign that though Chinese language was used to be considered a difficult language there has been a turnaround in this notion. At the moment we can say that what is difficult about learning Chinese as a foreign language is to achieve standard pronunciation. As we know, when two or more Chinese people from different provinces meet, their first encounter might be a problem, yet they are still able to interact with each other. It is apparent that standard pronunciation of Chinese language in itself is a theoretical baseline which can be achieved within acceptable range of deviation. Thus the assumption for the current study is that what matters is to be "heard", or comprehended by the listener.

We argue that, what is important is for the learner to be confident and be understood by the listener, but the question is how can the learner be confident for him or her to be understood? Confidence is one factor which is important in language learning, without it foreign language learners tend to produce the same word with varying tempo, varying pitch and varying tones thus in turn confusing the learner. We believe that if Chinese language learners' pronunciation is stable, even if their tone pronunciation might not be appropriate, then they will be still able to communicate with the native speakers of Chinese without any problem. Of course first encounter with a Chinese might be a problem but once the speaker and listener establish the pronunciation rule operating at that given moment there will be no communication challenges. Just as it would be when a Shandong person meets a Xi'an person, despite their differences in the dialects spoken, as soon as the two understand how each of them pronounce the same sound then they will be able to communicate and understand each other.

In order to build learners' confidence we therefore propose the use of cell phones or any other device which support voice recognition for Chinese language. Through use of voice recognition software, learners can establish whether or not their pronunciation is stable. In this paper we will demonstrate the use of voice recognition exercises where students can rehearse their speech production stability. The aim of the exercise is to establish whether the students can be able to consistently pronounce Chinese words and phrases which can be "understood" by the voice recognition device. If stable pronunciation which can be understood by the voice recognition device is achieved, learners can be confident that their pronunciation can be also understood by Chinese natives out there. Nevertheless, the use of cell phones as a tool for pronunciation practice is not anything new in the field of language learning; rather it is just the utilization of the existing technology in a way that helps to build learners' confidence.

4. Methodology

This study is designed as a software review's semi-experimental practical exercise. The experimental exercise was adopted mainly to raise consciousness on the use of the voice recognition application, which in some cases might be viewed as of little significance or too sophisticated for the teaching and learning of second language. In this paper we were concerned with whether or not Chinese as a foreign language learners at Hebei University were engaged by the researcher using the cell phone as the determiner of whether their pronunciation is or not appropriate. These exercises

follow the classroom pronunciation practice where the learner has to produce a given sound, word, phrase or sentence several times (Gilakjani, 2012). However, the differences between the classroom pronunciation exercise and this exercise is that when the cell phone replaces the role of the teacher, the evaluation of students' pronunciation is no longer verbal. Students will be able to read what their listeners hear from them. The student speaks, and the phone transcribes the words such that the student will be able to see where he or she is wrong. This method requires a quiet environment, stable internet and the speaker has to be as close to the phone as possible. At the end of the exercise we also used a questionnaire survey method to get feedback from the participants of this voice recognition exercise.

4.1 Apparatus used for the exercise

This exercise was done with a Lenovo p770 cell phone. Lenovo is a popular Chinese electronics brand which supports modern technology such as voice recognition, hence it supports Chinese voice input. The same exercise can be done with any other device as long it supports voice input. Because the Lenovo voice recognition software cannot operate without internet connection, we also used WIFI internet with standard speed of 10 Mbps.

4.2 Participants of the Exercise

Twenty students (female= 12; male=8) aged between 20 and 30 from different African countries who were studying Chinese as a foreign language at Hebei University at the time when this research was conducted participated in this voice recognition exercise. All the participants of this exercise had studied Chinese as a foreign language for at least three months and were able to read simple characters. The participants came from diverse first language backgrounds; for example, French speakers from Cameroon and Congo, Arab speakers from Sudan and Egypt, English speakers from Zambia, South Africa, Ghana, Liberia, Malawi and Zimbabwe as well as Portuguese and Spanish speakers from Mozambique and Madagascar respectively. This is typical of Chinese as a foreign language classes in China.

4.3 Procedures

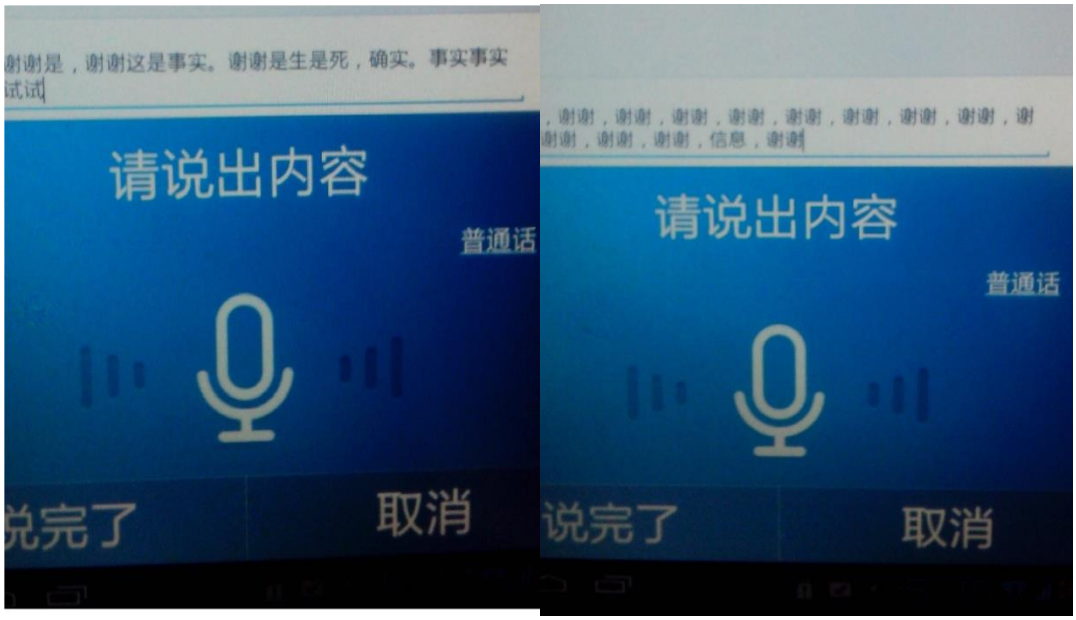
For the purpose of the current study, the words for the exercise were written down both in characters and pinyin as shown in Appendix 1. Each student was then given a chance to become familiarized with the words before the voice recognition icon on the phone was switched on. Students repeated the same word for at least five times, after which the phone would transcribe the words into Chinese characters. The transcribed words would therefore determine whether or not the student's pronunciation was correct. The stability of the student's pronunciation was determined by counting the number of correctly transcribed words versus the incorrect ones. The voice recognition software transcribes various words depending on one's pronunciation. Where the tone is wrong a closer word is transcribed and where both the tone and consonant is produced wrongly a completely different word is transcribed. In cases where the student's pronunciation was correct and stable, the correct word is repeated throughout. In some cases a correct word and an incorrect word are transcribed, showing that student's pronunciation is not stable. This is illustrated in the following section of the voice recognition exercise.

5. Results

Table 1 shows a list of 10 Chinese words which were produced by 20 students and various near correct pronunciation words were transcribed by the voice recognition software. For each word only four different variations were recorded for the purpose of this research, however during the real exercise more variations were noted per student.

As shown in Table 1, variations of words which recurred most frequently among the 20 students were presented in their order of closeness to the correct pronunciation where variation 1 represents the closest pronunciation, variation 2 representing a closer pronunciation, variation 3 representing a far deviation and variation 4 representing an extreme deviation. The standard deviation which can allow learners to be heard was defined as 1 and 2 while 3 and 4 were defined as far deviation which will distort the meaning of the word. The exercise showed that Chinese language learners from different countries can imitate Standard Chinese with different levels of accuracy, ranging from close deviation to extreme deviation. In some cases different students with different accents would produce the same word differently, yet the voice recognition software transcribed it as correct. From the above variations it can be concluded that though learners of Chinese are from different countries and despite their accents their pronunciation can still be understood. Apart from the above, participants were also able to determine whether their pronunciation was stable or not as shown in the following screen shots.

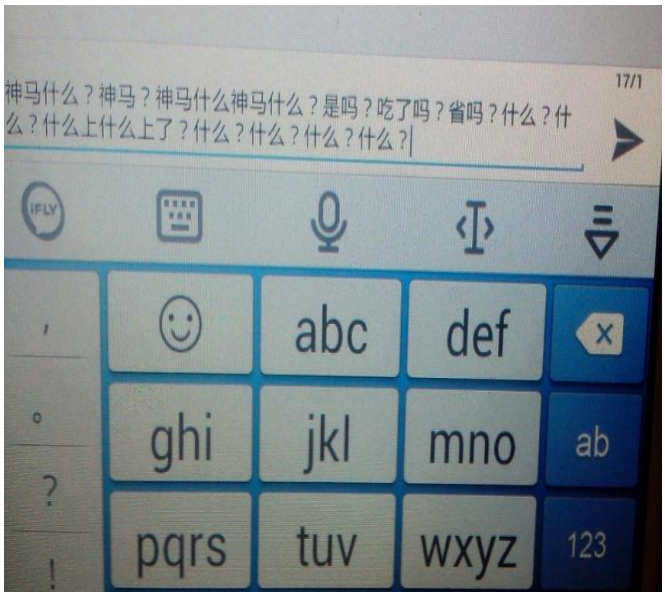
Figure 1. Screen Shots for the Word xiexie (谢谢)



The above two screen shots for the pronunciation of the word 谢谢(xiexie) are for two participants one of which had an unstable pronunciation and one with stable pronunciation. On the first screen shot the participant's pronunciation was not stable since at times the word was transcribed as "shengshi", "shisi", "queshi", 事实"shishi" and 试试"shishi". This clearly shows that this participant's pronunciation challenge is not on producing the word "xiexie", but on consistent production of the word. The second screen shot shows about 15 correct transcriptions of the word xiexie, this is evidence that the speaker is consistent in pronunciation. In such cases the learner will be able to go out there and confidently speak because he is sure that his pronunciation can be understood.

We also observed that for some words the more the learner continues to repeat the word the more he or she is able to produce the correct pronunciation. Apparently, the participants were more sensitive to the transcribed words such that when they saw wrong characters being transcribed they would adjust their pronunciation, as seen in the following screen capture of the word 什么"shenme".

Figure 2. Screen Capture of the Word shenme "什么"



In the above screen shot the first roll shows seven wrong pronunciation of the word "shenme" which were recorded as "Shenma", "shima" or "chima" but in the second roll eight correct words were transcribed out of eight. This suggests that when the learner makes an error, s/he is able to adjust the pronunciation even without the guidance of the teacher. The use of voice recognition software make language learners adjust their pronunciation when they notice the wrong transcriptions such that the more s/he produces the same word the more s/he will be able to produce it correctly. This is probably due to the fact that when a correct transcription is picked the speaker will try to maintain the same pronunciation.

In the above example, though the speaker was able to produce more correct words later but there were some times when wrong character would be transcribed. This shows that pronunciation practice is a long produce where the speaker experiments with different pronunciation strategies until s/he is able to stabilize the pronunciation in terms of tone, pitch and tempo.

The majority of students had problems in producing certain consonants, in such cases participants were able to see what their listeners are likely to hear. In the following example the participant produced the Chinese consonant /x/ as /sh/, so instead of producing "xiexie", closer words with /sh/ where transcribed such as "shishi", "shishui", "shishuai", for example.

Figure 3. Screen Capture for the Word xiexie“谢谢”

More screen shots for students' pronunciation showing pronunciation variations are listed in Appendix 1. In this exercise, talking to the phone was interesting because the phone would respond to what it hears, this is different from human to human conversations where sometimes the listener would pretend like they have understand or discourage the learner telling them that they do not understand anything.

After the pronunciation survey we administered a questionnaire to the 20 participants of this exercise. It was observed that all the participants had smart phones which had the voice recognition application; however, none of them had knowledge of using this application as a tool for practicing pronunciation. Although all of the participants could not correctly produce all the words for this exercise, they found this method interesting and worth using as a complimentary method for their pronunciation practice after class. During the practice 60% of the participants were able to recognize their errors. Because these participants had studied Chinese for three months thus some could not recognize Chinese characters for complicated characters. This implies that the exercise requires learners who can recognize Chinese characters. In this paper we suggest that those students who passed HSK level 2 and above can find this method useful since they can read at least 300 Chinese characters.

The majority of the participants showed interest in using this method on their own practice session since it is easy and convenient, however only 20% see this method as a complete substitution of the tradition pronunciation practice method. This indicates that at the moment voice recognition method is only useful as a complimentary method rather than a substitute of the learner–teacher practice pronunciation practice method.

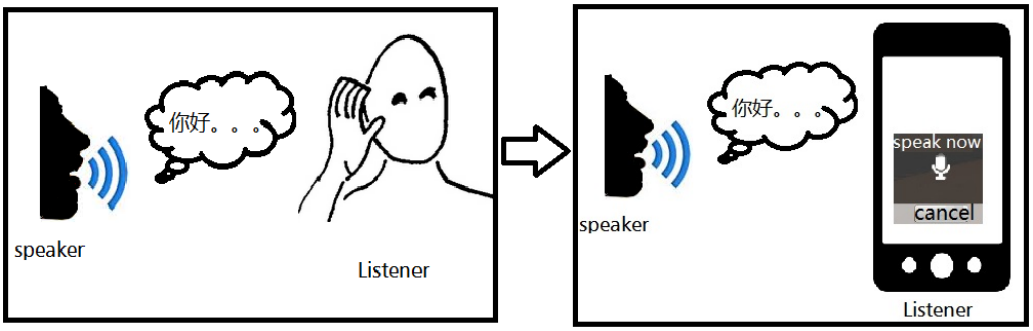
At the end of the exercise students were advised to use their phones as a tool to check their pronunciation. Voice recognition on cell phones is normally designed in such a way that even dialectical pronunciation can be transcribed as standard pronunciation, in other words, the voice recognition for Chinese phones is designed in such a way that close pronunciation would be understood. Hence, learners of Chinese as a foreign language should strive to at least achieve a level of pronunciation where at least the voice recognition applications can be able to transcribe their pronunciation correctly. The current study did not aim to prove a theory, but to demonstrate that

through the use of voice recognition students can be able to realize their pronunciation deviation and at the same time they can become aware of whether or not their pronunciation is or not stable.

6. Discussion

Use of voice recognition as an aid in practicing pronunciation is something that students can do as a rehearsal for the real human to human conversation or student to teacher pronunciation practice. Kettenhofen (2012) argues that rehearsing minimizes 75% of one's nervousness. A number of researchers agree to the notion that nervousness is one of the factors that causes second language learners not to attain high levels of proficiency (Aida, 1994; Gardner & MacIntyre, 1993; Horwitz, et al., 1986; Saito, et al., 1999). Therefore, with the use of the voice recognition pronunciation practice descibed in this article, one can easily overcome the nervousness which affects many other second language learners. To understand the role of the cell phone or any device with voice recognition application we can use the following two pictures for demonstration.

Figure 4. The Simple Model for Voice Recognition Pronunciation Practice



The emergence of voice recognition technology is therefore a transition from the human speaker and human listener interaction to human speaker and machine speech recognizer (machine listener) as shown above. Harper, et al. (2008) observe that modern mobile devices offer much more pleasurable human to mobile device interactive experience, this shows the gradual rise of the use of mobile devices and their substitution of human to human interaction. Instead of speaking to a person who will judge the pronunciation based on his or her personal perception of the sound, the cell phone becomes the decoder of the words or phrases based on the pre-coded standard deviation for pronunciation of different words. Based on the transcribed speech one can therefore determine whether or not his/her speech production can be understood by human listeners. When we speak a foreign language to another person we normally do not get to see what we are saying ourselves. In a student-teacher pronunciation practice which remains the best practice method, the listener who is the teacher will constantly correct the pronunciation. In daily conversations the listener will either show facial expressions showing that he or she did not understand or will ask “what? (什么?) , speak again (请再说一遍), I don’t understand（我不懂）etcetera. In all cases the listener is not able to see his/her pronunciation. However, as illustrated above the listener will be able to see whether his or her pronunciation is wrong.

Practicing Chinese pronunciation with the phone as the listener makes the environment amusing and interactive, even without a language practice partners. This is more of synchronizing learning with entertainment, a practice method which boasts interest in language learning because it is a kind of 'edutainment'. The Online English dictionary defines 'edutainment' as a method of learning through a medium such as computer software that both educates and entertains. This is not to suggest that edutainment is the best method of learning a language, neither are we proposing that

learning is an unpleasant experience, rather the argument here is that synchronizing the two helps the learners to learn while they are playing with their smart phones, that way learner's cell phone become a closer partner in language learning rather than an obstacle that distracts learning. As Ito (2009) puts it, edutainment promotes active participation and open ended learning. Some students are inactive in class due to various reasons such as; being shy, fear for making errors and time limitation in the classroom setting. However, with the use of voice recognition applications on the phones, students are able to actively practice their pronunciation anytime without any fear to be embarrassed before the whole class and at the same time they get immediate feedback. Voice recognition software does not comment on students' performance, but shows the student the type of error made. This kind of feedback helps especially those students who are not aware of their pronunciation abilities. In the classroom setting when the teacher gives negative feedback on a student's pronunciation, either the student can feel discouraged or the teacher will end up getting used to such an extent that even if errors are made s/he will not bother correcting them.

When students practice Chinese pronunciation, it is not how best they can produce a given single sound; rather what matters is whether or not their listeners can understand. In some cases students get discouraged when they fail to produce Chinese sounds as the native speakers of Chinese would produce them; however we believe that with the use of voice recognition application one can practice and master how to adjust his or her pronunciation so that listeners can understand despite the accent. Integrated use of voice recognition application with learning of Chinese is not the best method of learning Chinese pronunciation but a useful complementary method which can increase learners' confidence.

7. Conclusion

This paper attempted to demonstrate how voice recognition can be integrated into language learning. We conclude that while the traditional pronunciation practice still remains vital in foreign language learning, voice recognition applications on cell phones provides a broader platform for pronunciation practice. The voice recognition exercise which involved 20 Hebei University African students shows that technological development is no longer as complicated as it used to be, this gives language learners more opportunity to practice their pronunciation even without the teacher. Using this exercise, students no longer needed practice partners, because their cell phones acted as the listener who judges whether the learner's pronunciation can be heard or not. The exercise shows that learners can see the type of pronunciation error which they are making thus helping them to develop an appropriate pronunciation strategy which can be understood by human listeners.

This exercise helps learners to carry out a pronunciation practice which emphasizes making listeners comprehend. Because the practice can be done without a human language practice partner, the learner can double his or her pronunciation practice time, where voice recognition pronunciation practice using the phone act as a rehearsal of the real-life conversation. This way the learner will be more confident and more conscious of the principle of communication, that is, to convey information.

It is our hope that as technological advancements continues to improve the efficiency of voice recognition applications, language teachers and language learners will also continue to explore more ways of utilizing such applications. This will certainly ease both the teaching and learning of difficult foreign languages such as Chinese. However, the complimentary role that this pronunciation method plays in teaching foreign language still need further research with a more systematic analysis on students' reaction and ability to utilize the application.

The main challenge currently is that language teachers rarely teach their students how to utilize voice recognition for pronunciation practice. Integrating voice recognition with language learning is a potential teaching aid which can help students to have more confidence. One other challenge is that currently the voice recognition applications currently available require internet access, and this limits

the use of this exercise only to those who have access to free or cheap internet. We look forward to the development of more voice recognition applications that do not require access to the internet.

References

Abeysinghe, T. & Lu, D. (2003). China as an economic powerhouse: Implications on its neighbors. *China Economic Review*, 14; 164 – 185. Retrieved from http://courses.nus.edu.sg/course/ecstabey/Tilak-Ding.pdf.

Aida, Y. (1994). Examination of Horwitz, Horwitz, and Cope's construct of foreign language anxiety: The case of students of Japanese. *Modern Language Journal*, 78(2): 155-168.

Bada, E. (2001). Native language influence on the production of English sounds by Japanese learners. *The Reading Matrix*, (1)2.

Cook, V. (1991). *Second language learning and language teaching*. London: Edward Arnold.

Counselman, D. (2010). *Improving pronunciation instruction in the second language classroom*. Unpublished PHD dissertation. The Pennsylvania Utate University.

Cummins, J. (1984), *Bilingualism and special education: Issues in assessment and pedagogy*. Clevedon: Multilingual Matters.

Duanmu, S., (2007). *The Phonology of Standard Chinese* (2nd Edition). Oxford: Oxford University Press.

Furui, S. (1999). *Automatic Speech Recognition and Its Application to Information Extraction*. Proceedings of the 37th meeting of ACL, 11—20.

Gales, M. & Young, S. (2007).The Application of Hidden Markov Models in Speech Recognition. *Foundations and Trends in Signal Processing*, (1) 3; 195–304.

Gardner, R. C. & MacIntyre, P. D. (1993). A student's contribution to second language learning, part 11: Affective variables. *Language Teaching*, 26(1): 1-11.

Gilakjani, A. P. (2012).The Significance of Pronunciation in English Language Teaching. *English Language Teaching*, (5)4; 96-107, Retrieved from http://www.ccsenet.org/journal/index.php/elt/article/view/15940.

Harper, R., Rodden, T., Rogers, Y. and Sellen, A. (2008). *Being Human: Human-Computer Interaction in the year 2020*. England: Microsoft Research Ltd 7, Cambridge.

Horwitz, E. K. Horwitz, M. B. & Cope, J. A. (1986). Foreign language classroom anxiety. *Modern Language Journal*, 79(1): 125-132.

http://www.thefreedictionary.com/edutainment.

Hsin, E. & Hedge, A. (1999). *Survey of Continuous Speech Recognition Software Usability*. Cornell University, Retrieved from http://ergo.human.cornell.edu/ahprojects/hsin99/voice%20recognition%20paper.pdf.

Hyltenstam, K. & Abrahamsson, N. (2000). Who can become native-like in second language? All or some or none? *Studia Linguistica*, 54(2); 150-166.

Ito, M. (2009). *Engineering Play: A Cultural History of Children's Software*. Cambridge, MA: MIT Press.

Juang, B.H. & Rabiner, L.R. (2004). *Automatic Speech Recognition – A Brief History of the Technology Development*, Retrieved from http://www.ece.ucsb.edu.

Kettenhofen,C. (2012). Public Speaking Skills: The Importance of Rehearsing. *Bounce Back Higher*, Reterived from http://BounceBackHigher.com.

Laroy, C. (1995). *Pronunciation*. Oxford: Oxford University Press.

Levis, J. & LeVelle, K. (2011). Pronunciation and intelligibility: An overview of the conference. In. J. Levis & K. LeVelle (Eds.). *Proceedings of the 2 nd Pronunciation in Second Language Learning and Teaching Conference*, Sept. 2010. (pp. 1-6), Ames, IA: Iowa State University.

O'Hear, S. (2010). *Babbel introduces speech recognition to aid language learning*. Internet article retrieved from http://techcrunch.com/2010/06/23/babbel-introduces-speech-recognition-to-aid-language-learning/.

Rolandi, W. (2005). Speech Recognition in Education: Unexploited Opportunities. *Speech Technology Magazine*, Available on http://www.speechtechmag.com.

Rosenberg, A. (1979). The hardest natural languages. *Lingvistice Investigations.* (3)2; 323-339, Retrieved from http://people.cs.umass.edu/~rsnbrg/hardest.pdf.

Saito, Y., Horwitz, E. K. & Garza, T. J. (1999). Foreign language reading anxiety. *Modern Language Journal*, 83(2): 202-218.

Sangani, K. (2013). Voice recognition comes of age. *Engineering and Technology Magazine*. 8(7), Available on http://eandt.theiet.org/magazine/2013/07/inner-voice.cfm.

Scovel, T. (1988). *A Time to speak: A psycholinguistic inquiry into the critical period for human speech*. New York: Newbury House.

Shen, Y. (1958). *Learning the Chinese script can be easy language learning*, (8)1; 17-30. Retrieved from http://onlinelibrary.wiley.com/doi/10.1111/j.1467-1770.1958.tb01213.x/pdf.

Zechner, K., Bejar, I. & Hemat, R. (2007). Toward an understanding of the role of speech recognition in nonnative speech assessment. *TOEFL iBT Research Report,* http://www.ets.org/Media/Research/pdf/RR-07-02.pdf.

Appendix 1: Screen shots for the 10 words

赔偿peichan

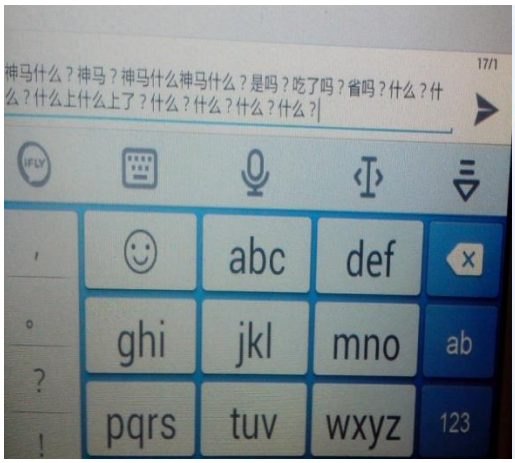
妈妈mama

喝水heshui

知识zhishi

吃饭chifan

亲戚qinqi



什么shenme

你好nihao

谢谢xiexie

搜Sou

Appendix 2: Questionnaire for use of voice recognition application

Answer the following questions with Yes or No, using **Y** for Yes and **N** for No.

1. Do you have a smart phone with voice recognition application?
2. Did you know about using voice recognition application for practicing pronunciation?
3. Were you able to realize your pronunciation errors during the voice recognition practice?
4. Would you prefer to use your phone for pronunciation practice rather than practicing with your teacher or friend?
5. Do you think voice recognition applications should be used in learning Chinese?
6. Did you find the practice session interesting?
7. Were you able to produce all the words correctly?
8. Does this practice method make you enjoy learning Chinese?
9. Will you use this pronunciation practice on your own?
10. Are you now able to use this method alone without somebody's help?

Table 1: Variations of Words for the 20 Students

	Variation 1	Variation 2	Variation 3	Variation 4
Right pronunciation	Closest pronunciation	Closer pronunciation	Far deviation	Extreme deviation
赔偿péicháng	陪唱péi chàng	非常fēi cháng	培养péi yǎng	北疆běi jiāng
妈妈mā mā	慢慢màn màn	蛮蛮mán mán	买卖mǎi mài	那么nà me
喝水hē shuǐ	和谁hé shuí	瞌睡kē shuì	黑水hēi shuǐ	也睡yě shuì
知识zhī shi	只是zhī shì	指示zhǐ shì	既是jì shì	即使jí shǐ
吃饭chī fàn	吃法chī fǎ	鸡饭jī fàn	师范shī fàn	日饭rì fàn
什么shén me	事实shìshi	沙马shā mǎ	神马hén mǎ	是吗shì ma
你好nǐ hǎo	你号nǐ hào	你啊nǐ ā	你收nǐ shōu	以后yǐ hòu
亲戚qīn qī	秦池qín chí	亲亲qīn qīn	其器qí qì	心情xīn qíng
谢谢xiè xiè	婶婶shěn shěn	色色sè sè	四岁sì suì	全省quán shěng
搜sōu	收shōu	瘦shòu	苏sū	松sōng

Table 2: Results of the Questionnaire Survey

	Percentage
Students who had cell phones	100%
Knowledge about voice recognition as a pronunciation practice	0%
Students who could realize their errors	60%
Students willing practice on their cell phones without teacher's help	20%
Those who think that voice recognition applications should be used in learning Chinese	80%
Number of students who found the method interesting	75%
Those who produced all the words correctly	0%
Those who think that this method make learning of Chinese interesting	65%
Those who are willing to use voice recognition for pronunciation practice as a complimentary method	100%
Those who could use Voice recognition method without somebody's help after the practice	100%