

APPLICATION OF THE SPEECH RECOGNITION  
TECHNOLOGY IN LANGUAGE EDUCATIONMehryar Nooriafshar<sup>a\*</sup>, Darius Nooriafshar<sup>b</sup><sup>a</sup>Faculty of Business, Education, Law and Arts, University of Southern Queensland, Toowoomba, 4350, Australia<sup>b</sup>University of Southern Queensland, Toowoomba, 4350, Australia**Abstract**

As technology advances and innovative communication products emerge, the interaction between people, globally, changes. Therefore, technology plays a major role in influencing both language and culture. The purpose of this paper is to answer the question of how the speech recognition technology available on the smartphones and tablets can support language learning. The very latest and advanced features of the smartphones such as the speech recognition capabilities were utilised in experimenting with language learning. The languages included French, German, Italian, Japanese and Mandarin. In all cases, the established learning concepts such as learning by guidance were considered. This paper has demonstrated how the latest technologies such as speech recognition can be utilised to produce effective educational materials for immersive language education. It is concluded that Technology can enable a learner to simulate the learning by guidance approach in the absence of the tutor or the opportunity of being in the actual environment.

*Keywords:* Language, learning by guidance, speech recognition

© 2016 Published by Future Academy. Peer-review under responsibility of Editor(s) or Guest Editor(s) of the EjMS.

\*Corresponding author.

E-mail address: Mehryar@usq.edu.au

doi: 10.15405/ejms(2421-8251).2016.1.7

**1. Introduction**

One needs to determine the Purpose of learning new languages before discussing the methods. Firstly, it must be emphasised that our world is becoming more connected all the time. People have become more interested and capable of communicating with each other. Of course, this communication is not principally face to face. Never-the-less, courtesy of the modern communication styles, it is a form of quick, easy and "24/7" style of interaction.



Different forms of technological advances in inexpensive, comfortable and faster means of transportation have also played a major role in connecting people.

Although the world is made up of different types of cultures, languages all stem from the same source - the human thought. In order to convey our thoughts to others orally, we would utilise sounds representing objects, creatures or our feelings. In its primitive form, the basis was an imitation of what represents the content of thoughts. Over the years these sounds have evolved into what we use now. Due to differences in geography and environmental conditions, different languages have taken different paths of development.

It would be logical to accept that a common language could unite the different people. The English language is accepted by the majority to be universal. Many people from the non-English speaking backgrounds have gained a basic appreciation of English. The English language's popular acceptance is due to a number of factors. Firstly, English is a relatively easy language to learn at an introductory level. Mastering English, however, takes many years. English is one of the only languages which has abandoned the communication barriers in a very interesting manner.

The connection between culture and language is strong and in most cases the influence is a two-way process. In other words, the language has an effect on the culture and culture affects and develop the language. Due to very specific features of English, the Anglo-Saxon culture has become very familiar to many non- English-speaking nations.

One of the main purposes of this paper is to explore and identify possible innovative educational products and the emerging technology's applications in education.

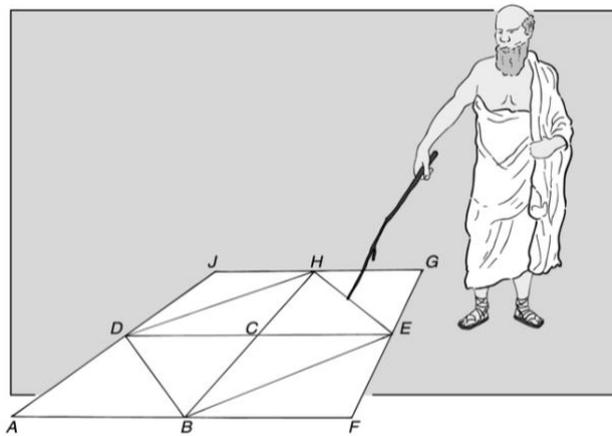
## **2. Traditional Learning and Teaching Approaches**

Learning and teaching approaches have certainly been influenced by the emerging technologies. The learning methods will change even more dramatically in the years to come. One thing, however, remains the same; and that is the ability of the teacher (human or machine) to convey the underlying concepts to the learner. Hence, the learner can build new meanings without simply memorising pieces of information received from the teacher. This way of learning is known as constructivism, which encourages the learner to construct their own meanings rather than simply memorising someone else's. Under constructivism the nature of learning takes a different form.

The underlying concept of constructivism goes back to the Socratic times. This way of learning encourages the learner to construct their own meanings rather than simply memorising someone else's. The concept of guiding and leading the learner to find out the solution or the right answer to a problem was discussed by Plato (the ancient scholar) almost 2400 years ago. If we analyse Plato's famous "dialogue" Meno, we will realise that Socrates demonstrates to Meno how a mathematically ignorant person solves a geometrical problem through a controlled guidance procedure rather than being told directly.

In the dialogue Socrates conducts his geometrical experiment on one of Meno's retainers who was totally ignorant of mathematics.

In this experiment, Socrates asks the boy to determine the dimensions of a square, which is exactly twice as large as a given square (say, abcd). The boy, eventually, after a series of questions, finds out that the correct solution is obtained by constructing the square (twice as large as abcd) on a diagonal (say, ac) of the given square. See Figure 1 for an illustration.



**Figure 1.** Socrates pointing to the Square  
(Source: The Author)

Even if learning is only the recovery of the pre-existent knowledge in the human soul, as Socrates argues, it can be passed on from teacher to learner by simply guiding the learner to find out for himself.

It should be emphasised that to start learning a language as a child, first of all we have listened, then imitated and finally spoken. Only after developing very basic conversational skills we have gradually learned about the rules governing the accepted usage of these sounds. Hence listening followed by speaking are the most effective ways of learning a language. We can adopt a similar child-like approach in acquiring the skills needed to have the basic conversational skills in a different language.

At the early stages of learning, ideally, only basic sentence structures consisting of only 2 to 3 words focusing on polite everyday expressions would be an effective approach. The widely used and available language instruction materials would be an ideal source for learning these basic phrases. Nowadays, these resources can be found as smart phone and tablet apps as well.

As observed by the author, Chomsky's idea of reliance on the biologically innate language faculty or "Universal Grammar" has re-emerged in modern language learning. For details, see Chomsky (1986). Some modern and innovative language teachers now believe that grammar should be learnt in an approach very similar to a child learning the grammatical principles. Michel Thomas, one of the innovative language teachers believed that the only grammar needed to learn a new language would be to understand what is meant by a verb, noun and adjective. He also adopted an approach in which one's development in learning the basic conversational skills of a language is like completing the framework of the building. In other words, the basic framework should be constructed first, and then, when necessary additions can be considered

At this stage, it would be appropriate to address technology in education, and in particular, language learning.

### **3. Technology Based Learning**

It is interesting to note that according to Moore's Law, the computing power is doubling almost every two years. It should be noted that the increase rate in the computing power, is subject to an exponential growth. It means that the doubling rate is also increasing not linearly but exponentially. On the other hand, the periods of two years is decreasing, most probably, at a negatively exponential rate.

So, what all this means is that we are having access to much more powerful computers every year. It is envisaged that by the year 2020, we will have a chip which will be as powerful and capable as a human brain. It is expected that by the year 2030, a computer chip which is 1000 times as powerful and capable as a human brain will be available. It should be noted that we are not just considering the number crunching capabilities of the computers. We are, however, expecting computers which will behave in a manner very similar to the human brain.

Hence it will not be inconceivable to have machines which will think and behave like human beings. It will be possible to have a meaningful dialogues with the computer. This idea is not quite far-fetched. New and emerging products are becoming available which have capabilities of creating natural interface with computers, the smart tablets and portable phones in the form of speech in several languages. For instance, this article has been written mainly by talking to a phone using the Dragon Dictation (Nuance, 2013) *app* which deciphers speech and then converts into text.

The author has also experimented practising other languages such as French, German, Mandarin and Japanese using this particular *app*. The capabilities of this particular app are amazing as it even has some learning abilities. For example, it learns the user's accent and also some of the frequently used words and nouns are remembered for future reference.

The following is an example of dictating a phrase in Japanese which the system has converted that to appropriate Hiragana, Katakana and Kanji.

「こんにちは。わたしのは、マリオです。あなたのはですか。 Konnichi wa. Watashi no namae ha, Mario desu. Anata no namae ha nan desuka.」

(Hello. My name is Mario. What is your name?)

Speech-to-Text Recognition (STR) technology synchronously transcribes text streams from speech input and displays them on a screen (Alapetite et al. 2009). It should be noted that modern smartphones such as iPhone have excellent features for converting speech into text. By changing the keyboard option into the language of choice, one can simply speak to the microphone and let the Smart phone or the tablet convert the speech into the text in that particular language. Hence, one would be in a position to experiment speaking and then seeing the text here on the screen. As suggested by Hwang et al. (2012), STR generated text can greatly assist an individual to attain a better understanding.

The text to speech features of these clever devices allow the user to playback what was recorded in writing. The playback will enable the user to see what was actually written. Therefore, one can, based on the feedback, adjust the pronunciation until the desired outcome is achieved. The following is an example of the recorded after speaking into the Mandarin people.

请问

我要這個

超市在哪裡

谢谢

With the above example, the author had pronounced the following:

Qing wen (Excuse me)

Wo yao zhe ge (I want this/this one)

Chaoshi zai nali (Where is the supermarket)

Xie xie (Thank you)

With a little bit of practice one can, easily, achieve the correct pronunciation for various phrases. Therefore, in the absence of an opportunity to be in an environment or have a tutor to correct one's pronunciation one can still have the immersion experience. The general approach to this kind of learning is very similar to the learning by guidance as in Meno.

The technology which has facilitated this kind of learning is progressing continually and more efficient products are being developed. Based on several years of research in the use of technology in education, the author foresees very interesting and practical learning and teaching applications for the emerging technologies.

#### 4. Conclusions

It was demonstrated that the concept of learning by guidance can be a very effective approach in general learning. The idea can also be applied to language learning. It was also concluded that the language should, ideally, be learnt in a natural manner. This approach is very similar to how one learns to speak as a child. Therefore, one should commence by listening first, and then imitating and speaking.

This paper has demonstrated how the latest technologies such as speech recognition can be utilised to produce effective educational materials for immersive language education. Hence, the technology enables a learner to simulate the learning by guidance situation in the absence of the tutor or an opportunity of being in the actual environment.

#### Acknowledgements

The author(s) declare that there is no conflict of interest.

#### References

- Alapetite, A., Anderson, H. B., & Hertzum, M. (2009). Acceptance of speech recognition by physicians: A survey of expectations, experiences, and social influences. *International Journal of Human-Computer Studies*, 67(1), 36-49. <https://doi.org/10.1016/j.ijhcs.2008.08.004>
- Chomsky, N. (1986). *Knowledge of Language: its nature, origin, and use*. New York, NY: Praeger.
- Hwang, W., Shadiev, R., Kuo, T. T., & Chen, N. (2012). Effects of Speech-to-Text Recognition Application on Learning Performance in Synchronous Cyber Classrooms. *Journal of Educational Technology & Society*, 15(1), 367-380.
- Nuance Communications. (2013), Dragon Dictation (Version 2.0.28) [iPad/iPhone app].