International Conference
ICT for Language Learning

Automated Speech Recognition: its Impact on Teaching and Learning Languages

Michael Carrier[1]

Abstract

Speech recognition has reached a level of accuracy where it is powering automatic translation and testing. What impact has it on language teaching? How should we develop appropriate pedagogical models and prepare teachers for its application to our classrooms? I will give a critical analysis of its pedagogical uses and dangers.

The technology of Automated Speech Recognition (ASR) is rapidly becoming more sophisticated and it is becoming part of everyday life. Earlier applications to language teaching were very inaccurate and did not aid learning. But the current accuracy levels of ASR mean that this is changing.

Speech recognition facilitates automatic translation. There are already mobile apps that allow students to speak into a phone or tablet and instantly hear the spoken translation. These 'speech-to-speech' systems are mainly accurate in narrow domains (eg domestic or tourist language) but are likely to impact on students' motivation and expectations of learning languages.

ASR facilitates computer-based automatic marking of language teaching examinations - both written and spoken exams. Cambridge University has set up a new institute, ALTA, to research this and is trialling the automatic marking of Cambridge English language exams.

It also facilitates auto-response to communicative interactions in the classroom, where students can use their tablets (in pairs) to speak or write responses to a task and get an instant correction or formative assessment. It also facilitates new ways to work on phonology and accent - using IBM's programme 'Reading Companion', for example.

I will give a critical appraisal of the application of Speech Recognition to language teaching and consider the impact on pedagogy and teacher development needs this may entail.

1. Introduction

The aim of this paper is to outline the impact on English language teaching of the use of automated speech recognition (ASR) technology.

In this paper I will discuss the nature of ASR, how it works and how it can be used in class.

I will also touch on the technology of speech to speech translation, which uses ASR and speech synthesis, and outline how this may also impact on student motivation and teacher course design.

Finally I will address how ASR can be used to automate language assessment.

2. What is ASR?

Automated Speech Recognition (ASR) converts audio streams of speech into written text. ASR is still imperfect but improving rapidly, and is based on big data – searching language corpora and finding matching patterns in data.

ASR is developing rapidly in terms of accuracy and is being used extensively in commercial applications. It is now available for educational use where its application to language learning is extremely relevant to modern digital learning practices.

ASR has a chequered history in ELT, and there have been many inadequate commercial English language learning software products which promised more than they delivered. With the new levels of accuracy and recognition quality, however, new opportunities arise.

3. How does it work?

ASR turns speech into written text by using a 'speech recognition engine'. Speech recognition engines need these components:

an acoustic model: which is takes audio recordings of speech and their transcriptions (taken from a speech corpus), and 'compiles' them into a statistical representations of the sounds that make up each word

[1] Highdale Consulting, United Kingdom

a language model or grammar file: A language model is a file containing the probabilities of sequences of words. A grammar file is a much smaller file containing sets of predefined combinations of words.
The ASR process can be illustrated thus:

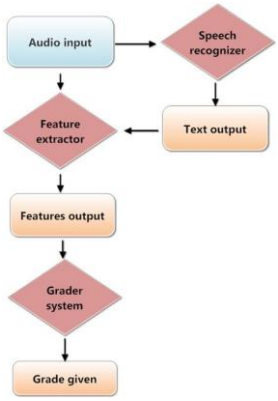

Figure 1

Once the spoken language has been recognised and turned into text, then the text can be analysed for meaning, and in the case of speech to speech translation turned into the text of the target language.

How Siri works

Many professionals will be familiar with speech-enabled services such as Siri, on the iPhone, or Cortana, on Microsoft phones. These services can respond to the user's spoken request or command. Understanding how they function is a guide to how ASR in ELT activities can also be used:
1 - The sounds of your speech are converted into digital data
2 - The data from your phone is sent to a server in the cloud which has an ASR engine
3 - The server compares your speech against a statistical model to estimate, based on the sounds you spoke, what letters the speech contains.
4 - Based on these opinions, your speech -- now understood as a series of vowels and consonants -- is then run through a language model, which estimates the words that your speech is comprised of.
5 - The computer then creates a list of interpretations for what the sequence of words in your speech might mean.
6 - The computer determines that your intention is, for example, to send an SMS, that person X is your addressee, that therefore her contact information should be looked up in contact list, and that the rest of the text is your actual message to be sent.
Then, your text message magically appears on screen ready to send.

4. Applications of ASR

ASR is already being used in various commercial sectors including telephone answering systems ('Tell us what you are calling the bank about') , in-car systems like navigation, military, healthcare, education, and different areas of disability support.
Applications include medical reports, created by doctors dictating their medical notes and having the transcription sent to other doctors without the usual timelag required for secretarial staff to transcribe spoken recordings. This saves time, money and potentially supports better medical outcomes.
Applications in the home include speech-enabled televisions, which allow users to simply speak instructions such as ‘please change the channel to BBC’ or ‘pause the DVD’ without using remote controls.
Computer and smartphone makers have included speech enabled services in all of their new systems, including Siri for the Apple phones or Cortana for Microsoft phones and computers.
Google has created a voice search which allows users to ask their search questions out loud and get answers spoken back whether you are out and to them.
Standard software such as Dragon Dictate allows anyone to create reports, memos etc by simply talking to their phone or computer.

5. How can we use ASR in ELT?

We can use speech recognition to support teaching, learning, writing skills, and assessment.
ASR can be used for a variety of activities, such as dictation, voice search on the web, pronunciation practice, translation practice, and marking of student production.
ASR facilitates new ways to work on phonology and accent – e.g. using IBM's programme 'Reading Companion', which listens to the student's pronunciation and corrects it by comparing to the stored pronunciations it expects to hear. IBM provides this speech-enabled literacy skill development free for students around the world.
ASR facilitates responses to communicative interactions in the classroom, where students can use their tablets (in pairs) to speak or write responses to a task, and get an instant correction or formative assessment.
A phone app called Speaking Pal provides activities where the learner interacts with a programme called English Tutor in short, real-life dialogues where the user controls the conversation flow, like in a real video / phone call. English Tutor is able to provide instant feedback on the student's speaking performance along with a review mode for later practice.
ASR can be used for storytelling tasks. Students tell a story by dictating to their device. One student takes the dictating role to ensure user accuracy. The group edits the resulting text and checks accuracy / appropriateness using online dictionaries.
ASR can be used for Conversation tasks, where for example students in pairs write a dialogue, perform it as dictation, then read and correct the written output.
ASR can also be used to get diagnostic feedback on Pronunciation issues
For self-study use, speech recognition can help learners to engage in speaking without a partner. The student works alone to dictate a text or act out a dialogue given by the teacher, dictating to a device. The ASR system provides written output which the student can then review and correct and share with the teacher for review.

6. Speech-to-Speech translation

Speech to speech translation seems like science fiction, like a reference to the Babel Fish (in Douglas Adams' book Hitchhikers' Guide to the Galaxy) which allowed aliens to speak to each other across the universe.
It means simply that you can speak to a device in your own language and it will automatically understand you, translate the message into another language, and immediately speak that message out loud to the conversation partner, in their own language.
It is used in products and services like Google Translate, the Apple Watch and Skype Translate.

How does it work?

All these systems use a machine translation system, which essentially looks for patterns in hundreds of millions of documents to help decide on the best translation, using a corpus of documents that have already been translated between the pair of languages by human translators. Finding patterns in this corpus helps the machine translation to make a statistically-informed intelligent guess about the best translation.
Of course this cannot be as accurate as a human translator, as it is looking at patterns in documents, not analysing and understanding meaning. But it can often provide a useful first translation.
This machine translation is linked up to a speech recognition system that listens and turns the user's speech into text - that can be input into the machine translation. Then the output, as text, is fed into a speech synthesiser, which speaks the result out loud.
Microsoft's Skype translate provides a variation on speech to speech translation and allows anyone to make a video or voice call with the person who speaks another language, and carry on a two-way conversation in both languages at the same time. At the same time an on-screen transcript of your call is displayed so you can follow the conversation easily.

7. Using speech to speech translation in class

A simple activity is for students to carry on a conversation using Google Translate but making notes of the nature of the translation choices made by the programme, to show discrepancies or to discuss alternative translations.

- Using Google Translate, students write a sentence or short text in L1
- Student A translates it into L2 (in English)
- Student B then speaks the English text into Google Translate and hears the L1 translation spoken by the device.

The two students then compare the versions of the text, and note differences created by the translation both ways, asking for teacher guidance where needed.

8. Automated assessment: marking students' speech output

Marking tests and giving individual feedback is one of the most time-consuming tasks that a teacher can face. ASR technology can be used to automate the marking and grading of students' spoken conversation. Speech produced in assessment situations can be fed through the same speech recognition process as described above, and in this case compared to language produced by students at different levels of the Common European Framework of Reference (CEFR).

> "Huge advances in speech recognition and machine learning mean that computers can now complement the work of human assessors, giving surprisingly accurate evaluations of language and helping to diagnose areas for improvement. " [1]

This can provide the benefit of speed - the spoken work from students does not need to be recorded and sent to a human examiner, but can be graded instantly.

> "Automated assessment won't replace human examiners anytime soon, but it can add great value to their work." [2]

9. Future trends in Speech Recognition for language education

It is clear that the growing use of portable and wearable devices such as watches and even glasses will lead to more ubiquitous use of speech recognition.

Human beings are speech-driven, and anything which allows them to carry out daily tasks using phones and computers without unnecessary typing will be extremely popular, both in commercial and educational applications.

One can predict the emergence of some of the future applications of ASR, which only a few years ago would have been considered science fiction yet are now being rolled out as products.

Speech to printed output

People still rely upon written documents, and it will soon be possible for people to dictate a letter, language activity response or creative composition simply by speaking and having this content immediately printed onto hard copy by a standard computer printer situated nearby, without any intermediate typing. This will help students with lower literacy and keyboard skills..

Speech activated equipment

People have to engage interact with equipment and devices, switching them on and off, changing TV channels, choosing washing machine temperatures et cetera. All of this will be speech enabled in the very near future and machines will respond to spoken commands - some televisions are already being sold with this feature installed.

Widespread automatic marking of speech

The grading of speech is likely to continue to develop very rapidly. For many less advantaged students, with lower literacy and writing skills, speaking is easier than writing an exam paper, and assessment carried out by automated speech recognition may well be fairer to these types of student than traditional examinations would be.

Speech recognition in education, and specifically language education, is here to stay.

References

[1] Retrieved from: http://alta.cambridgeenglish.org/
[2] Retrieved from: http://indiaeducationdiary.in/Shownews.asp?newsid=22241