CAMBRIDGE ENGLISH
Language Assessment
Part of the University of Cambridge

Speech Recognition in ELT: the impact on teachers and students

Michael Carrier

IATEFL Harrogate, 2014

CAMBRIDGE ENGLISH
Language Assessment
Part of the University of Cambridge

Contents

CAMBRIDGE ENGLISH
Language Assessment
Part of the University of Cambridge

1-What is ASR? What is it not?

- Automated Speech Recognition (ASR) converts audio streams into text, but does not analyse it semantically.
- The ASR output cannot assess meaning or coherence
- ASR is not the same as Natural Language Processing
- ASR is flawed but improving rapidly
- ASR is based on corpora and finding matching patterns in data

Speech recognition has come of age. It is accurate and part of everyday life, and powering automatic translation and testing systems.

What impact will this have on ELT and how should we develop appropriate pedagogical model, and prepare teachers for the application of speech recognition to our classrooms?

CAMBRIDGE ENGLISH
Language Assessment
Part of the University of Cambridge

ASR & ELT

- History of failure.
- ASR facilitates auto-re
interactions in the clas
use their tablets (in pai
responses to a task an
formative assessment.
- ASR also facilitates ne
and accent - using IBM
Companion' for examp
- Automatic translation.
that allow students to s
and instantly hear the
- These 'speech-to-spee
accurate in narrow domains (eg domestic or tourist language) but are likely to impact on students' motivation and expectations of learning English.
- ASR facilitates computer-based automatic marking of ELT examinations - both written and spoken exams. Cambridge University has set up a new institute, ALTA, to research this and is trialling auto-marking Cambridge ELT exams.

Small vocabulary / many-users
These systems are ideal for automated telephone answering. The users can speak with a great deal of variation in accent and speech patterns, and the system will still understand them most of the time.

However, usage is limited to a small number of predetermined commands and inputs, such as basic menu options or numbers.

Large vocabulary / limited-users
These systems work best in a business environment where a small number of users will work with the program.

While these systems work with a good degree of accuracy (85 percent or higher with an expert user) and have vocabularies in the tens of thousands of words, you must train them to work best with a small number of primary users. The accuracy rate will fall drastically with any other user.

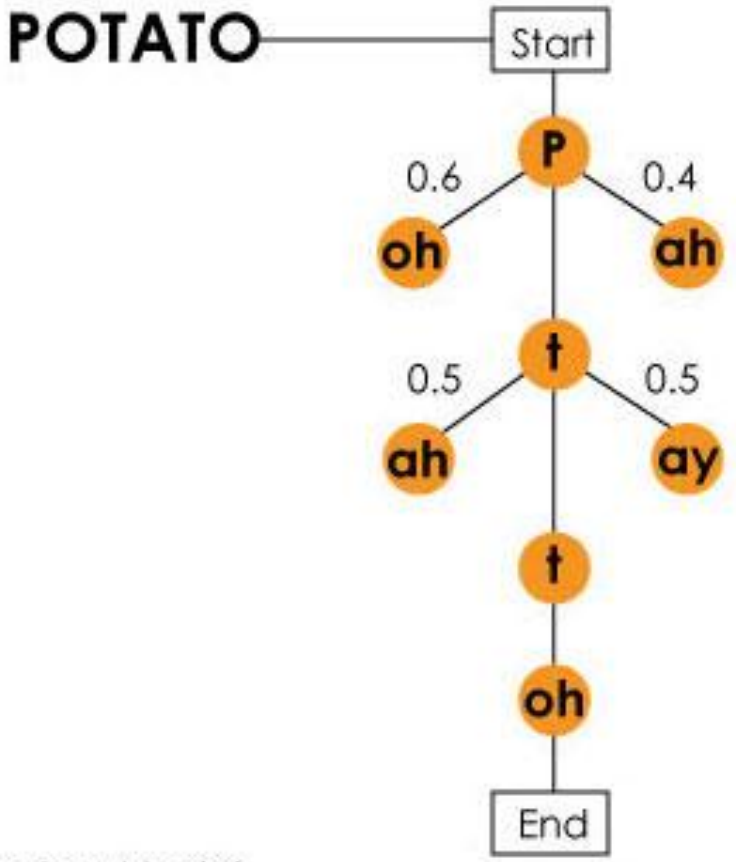
- **Speech recognition**, also referred to as **speech-to-text** or **voice recognition**, is technology that recognizes speech, allowing voice to serve as the "main interface between the human and the computer".

- Voice recognition can refer to products that need to be **trained to recognize a specific voice**, or those products used in automated call centers that are capable of recognizing **a limited vocabulary from any user**.

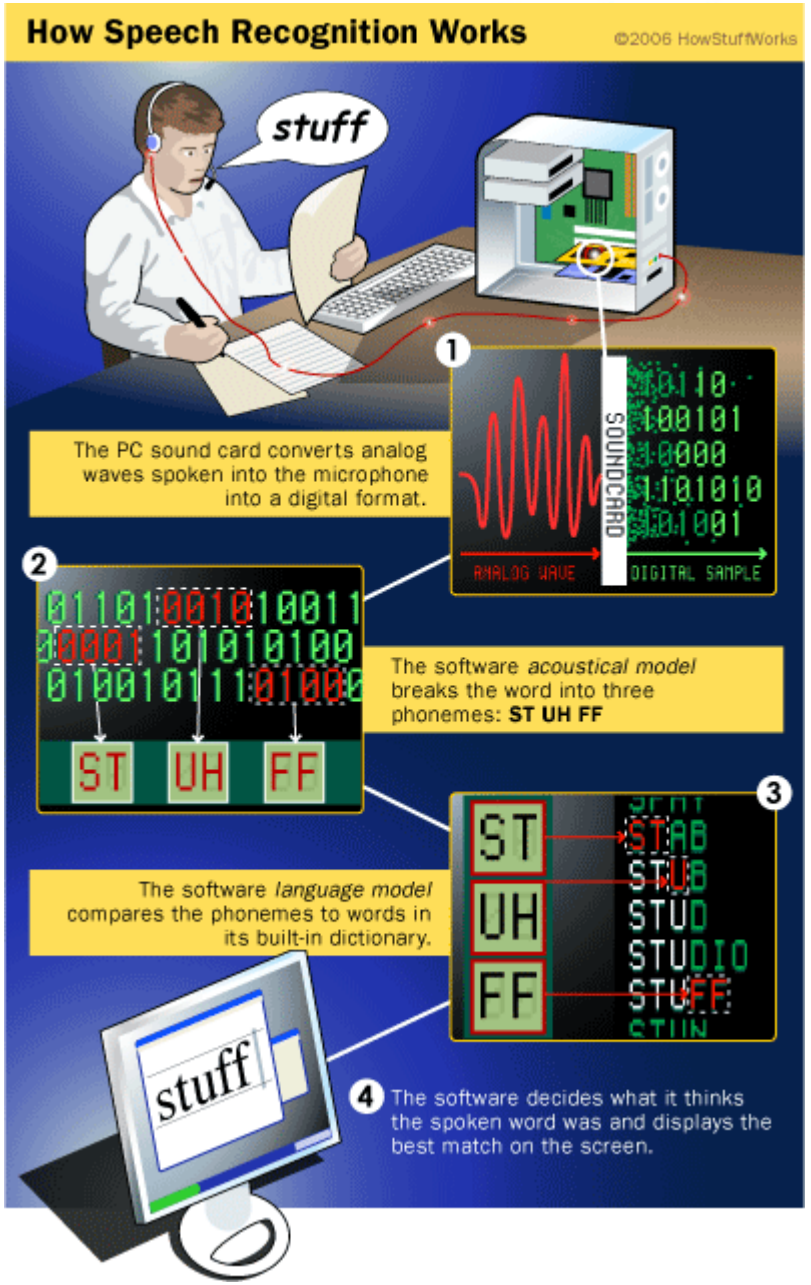
CAMBRIDGE ENGLISH
Language Assessment
Part of the University of Cambridge

2-How does it work?

How Speech Recognition Works

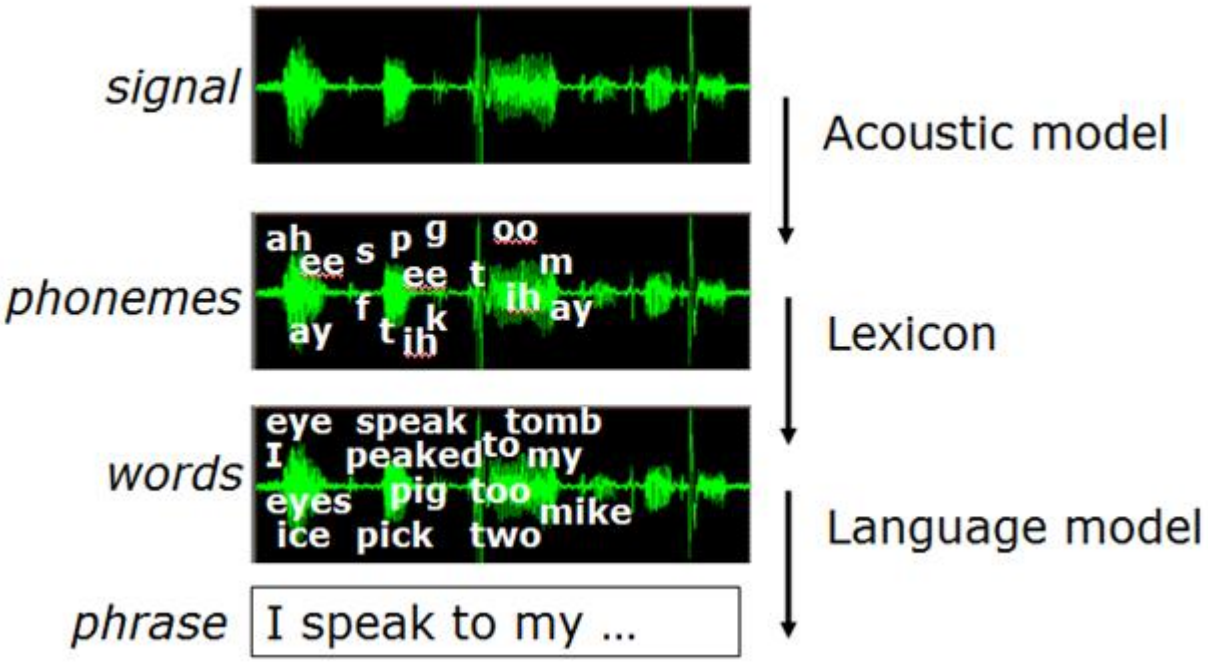


Markov Model



CAMBRIDGE ENGLISH
Language Assessment
Part of the University of Cambridge

2-How does it work?



program.

While these systems work with a good degree of accuracy (85 percent or higher with an expert user) and have vocabularies in the tens of thousands of words, you must train them to work best with a small number of primary users. The accuracy rate will fall drastically with any other user.

Speech recognition engines require:

an acoustic model, which is created by taking audio recordings of speech and their transcriptions (taken from a speech corpus), and 'compiling' them into a statistical representations of the sounds that make up each word (through a process called 'training').

a language model or grammar file. A language model is a file containing the probabilities of sequences of words.

A grammar is a much smaller file containing sets of predefined combinations of words. Language models are used for dictation applications, whereas grammars are used in desktop command and control or telephony interactive voice response (IVR) type applications.

Works?

Markov models

Vocabulary base

Corpora

Language modelling

Context dependency

Accuracy criteria:

- Vocabulary size and confusability
- Speaker dependence vs. independence
- Isolated, discontinuous, or continuous speech
- Task and language constraints
- Read vs. spontaneous speech
- Adverse conditions

Part-of-speech tags used:

MD	modal auxiliary (can, should, will)
NC	cited word (hyphenated after regular tag)
NN	singular or mass noun
NN$	possessive singular noun
NNS	plural noun
NNS$	possessive plural noun
NP	proper noun or part of name phrase
NP$	possessive proper noun
NPS	plural proper noun
NPS$	possessive plural proper noun
NR	adverbial noun (home, today, west)
OD	ordinal numeral (first, 2nd)
PN	nominal pronoun (everybody, nothing)
PN$	possessive nominal pronoun
PP$	possessive personal pronoun (my, our)
PP$$	second (nominal) possessive pronoun (mine, ours)
PPL	singular reflexive/intensive personal pronoun (myself)
PPLS	plural reflexive/intensive personal pronoun (ourselves)
PPO	objective personal pronoun (me, him, it, them)

CAMBRIDGE ENGLISH
Language Assessment
Part of the University of Cambridge

Siri

SS Activity – correct Siri

CAMBRIDGE ENGLISH
Language Assessment
Part of the University of Cambridge

How Siri works…

1 - The sounds of your speech were immediately encoded into a compact digital form that preserves its information.

2 - The signal from your connected phone was relayed wirelessly through a nearby cell tower and back to your Internet Service Provider where it communicated with a server in the cloud, loaded with a series of models honed to comprehend language.

3 - Simultaneously, your speech was evaluated locally, on your device. A **recognizer** installed on your phone communicates with that server in the cloud to gauge whether the command can be best handled locally -- such as if you had asked it to play a song on your phone -- or if it must connect to the network for further assistance. (If the local recognizer deems its model sufficient to process your speech, it tells the server in the cloud that it is no longer needed: "Thanks very much, we're OK here.")

4 - The server compares your speech against a **statistical model** to estimate, based on the sounds you spoke and the order in which you spoke them, what letters might constitute it. (At the same time, the local recognizer compares your speech to an abridged version of that statistical model.) For both, the highest-probability estimates get the go-ahead.

5 - 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. Given a sufficient level of confidence, the computer then creates a candidate list of interpretations for what the sequence of words in your speech might mean.

6 - If there is enough confidence in this result, the computer determines that your intent is to send an SMS, Erica Olssen is your addressee (and therefore her contact information should be pulled from your phone's contact list) and the rest is your actual note to her -- your text message magically appears on screen, no hands necessary.

Reflection

What is the impact of this for teachers in the classroom? What is the impact on teachers need for training and development to be able to use this technology in the classroom and adapt to its use in examinations?

3-How is it being used? Applications of ASR

Dictation

Voice search

Pronunciation

Translation

- Telephony
- In-car systems
- Military
- Healthcare
- Education
- Disability support – vision-impaired, RSI etc

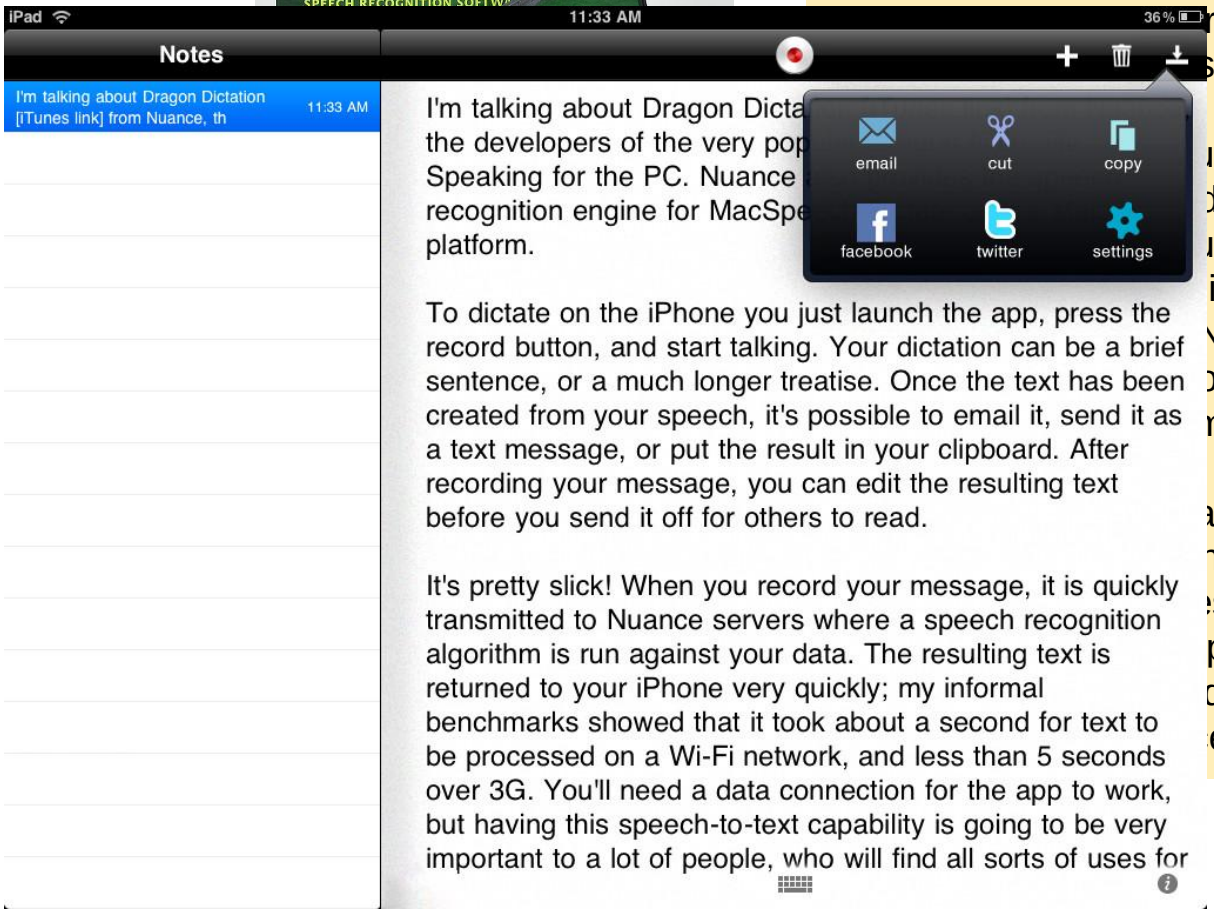
CAMBRIDGE ENGLISH
Language Assessment
Part of the University of Cambridge

Dragon

Pros and cons

Nuance today announced that Samsung's new GALAXY Gear wearable device and Samsung GALAXY Note 3 integrate Nuance's voice and language capabilities as part of Samsung's expanding lineup of S-Voice powered devices. Today's announcement also marks the first use of Nuance's voice and intelligent systems-based technology into the wearables
rger expansion of
s.

uance's voice
dsets, tablets, TVs
uance's voice
incredibly intuitive
Nuance has been at
onizing devices to
ns through voice,
technologies that
ay we access our
nce and Samsung
ess and
perience for
ds, learns and
es of the consumer.

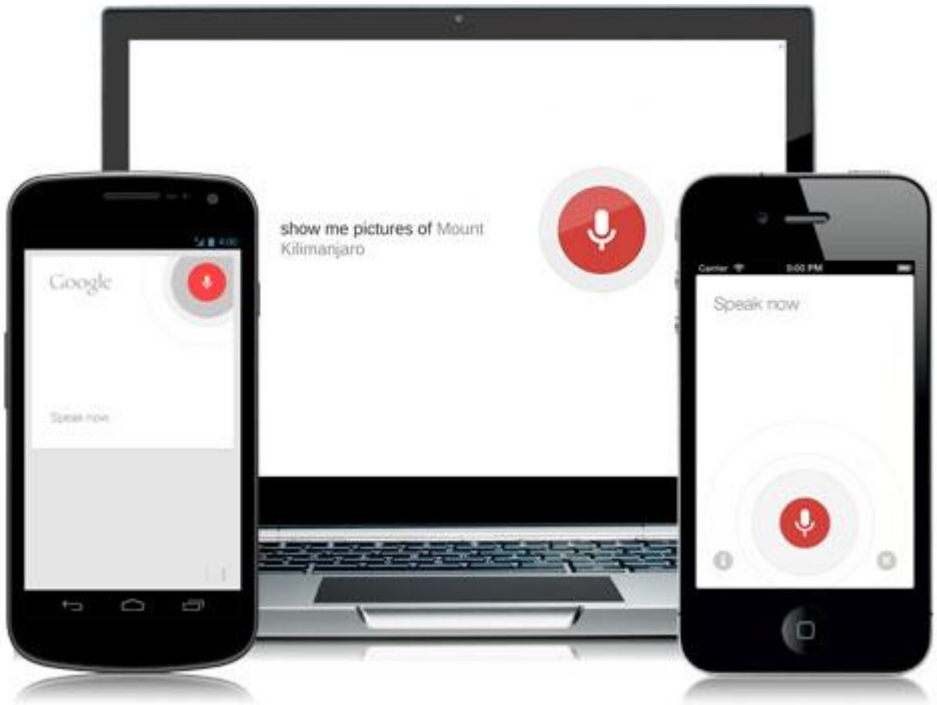


Activity

Use IOS

Google Voice Search

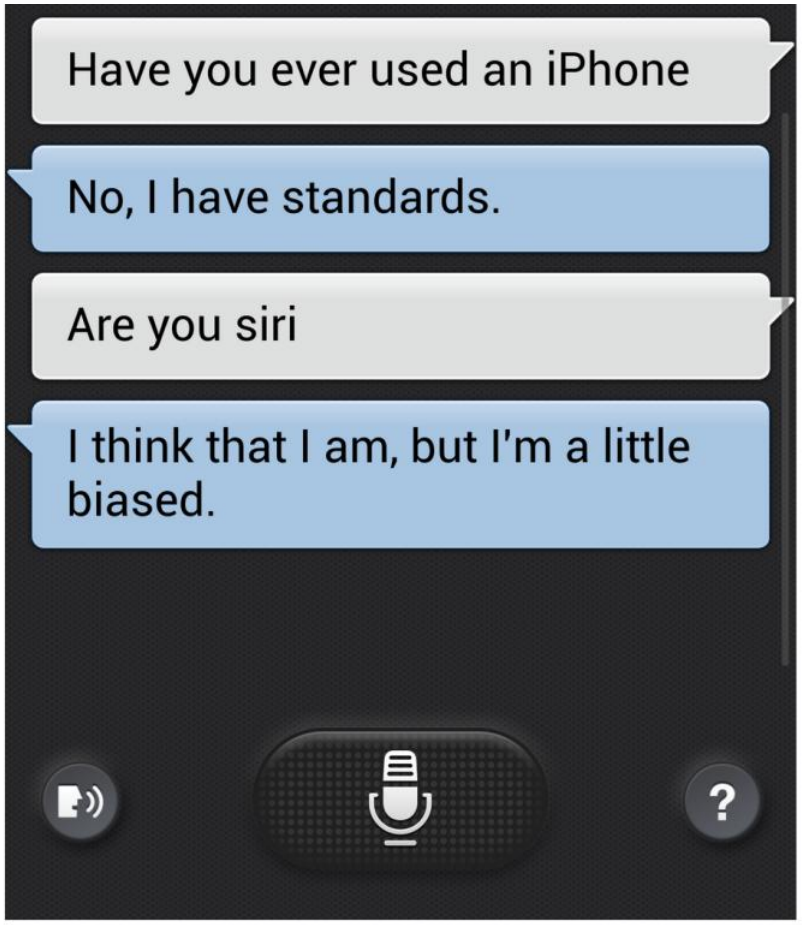
- Ask your questions out loud and get answers spoken back whether you are out and about or sitting at your desk. Just tap the mic on the Google search bar and speak up. This works on the Google Search App for iOS, Android and Chrome browsers for laptops and desktops.



Other ASR apps

Not just Siri…

Google Voice Search
Google Voice Typing
Vlingo
Nuance's Dragon Go!
True Knowledge's Evi voice assistant
Samsung S Voice
Microsoft's TellMe
Android's Speaktoit



CAMBRIDGE ENGLISH
Language Assessment
Part of the University of Cambridge

Knowledge Graph

Conversational Search:

Singhal stated, "A computer you can talk to? And it will answer everything you ask it? Little did I know, I would grow up to become the person responsible for building my dream for the entire world." Conversational search technology was then featured and Singhal introduced the term "hot-wording" to describe search without the need for an interface, whereby the user simply prompts the Google search engine by stating, "OK Google."

The I/O audience was then shown a demonstration in which a user asked a question about Santa Cruz and the search engine answered back in "conversation," in addition to the presentation of results for the query. Google's Johanna Wright explained that the search engine uses data from the Knowledge Graph to produce results: "The Knowledge Graph knows that Santa Cruz is a place, and that this list of places are related to Santa Cruz".

The **Knowledge Graph** is a knowledge base used by Google to enhance its search engine's search results with semantic-search information gathered from a wide variety of sources.

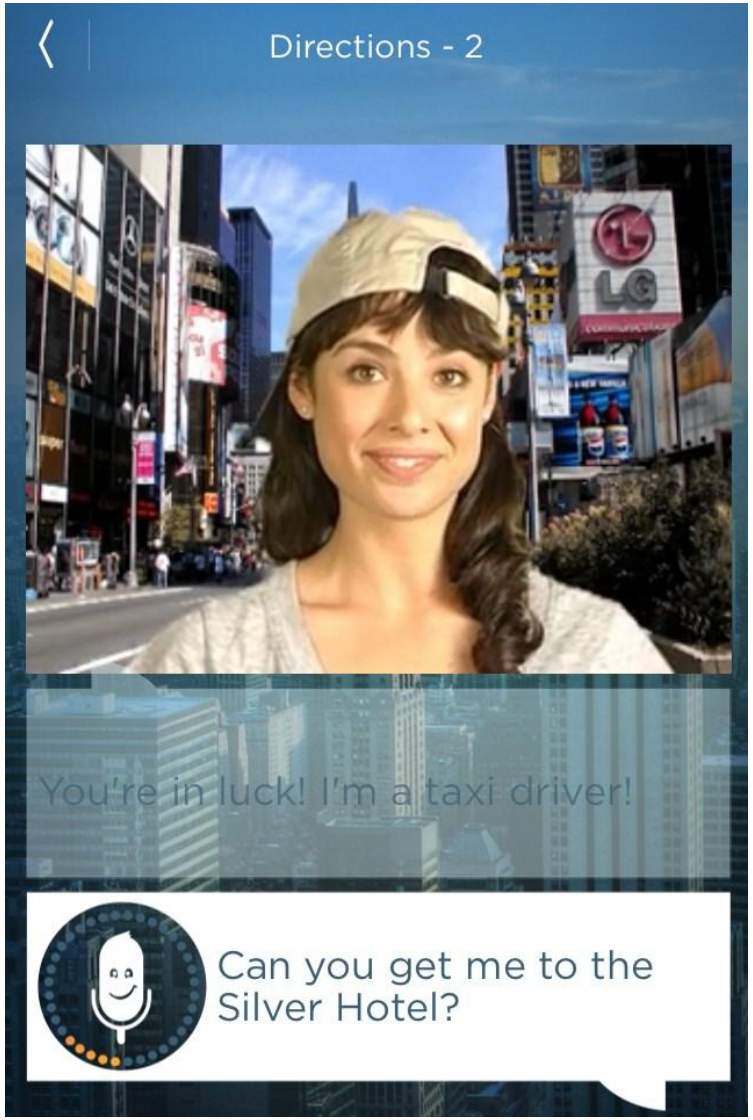
It provides structured and detailed information about the topic in addition to a list of links to other sites.

The goal is that users would be able to use this information to resolve their query without having to navigate to other sites and assemble the information themselves.

CAMBRIDGE ENGLISH
Language Assessment
Part of the University of Cambridge

SpeakingPal

Mini-lessons enable students to learn English in small sections that last 5 minutes or less (micro-learning). This unique methodology allows students to **learn effortlessly** during their daily activities while taking advantage of their idle time. The learner interacts with English Tutor in short, real-life dialogs where the user controls the conversation flow, like in a **real mobile video call**. Using SRI's state-of-the-art speech recognition technology, English Tutor is able to provide **instant feedback** on the student's speaking performance along with a review mode for later practice.



CAMBRIDGE ENGLISH
Language Assessment
Part of the University of Cambridge

IBM Reading Companion

Reading Companion is IBM's web-based literacy program that uses novel speech-recognition technology to help adults and children gain and increase literacy skills. Reading Companion's innovative software "listens" and provides feedback, enabling emerging readers to practice reading and pronunciation as they acquire fundamental reading skills.

How does it work?

Basically, users log on to the Reading Companion web site and are presented with material to read. An on-screen mentor, or companion, "reads" a phrase to the user and then "listens" to the user read the material through a headset microphone. Based on what was heard, the companion either provides positive reinforcement (e.g., "You sound great!") or an opportunity for the user to try reading a word again. As the user's skill improves, the technology reads less material so that the learner reads more.

CAMBRIDGE ENGLISH
Language Assessment
Part of the University of Cambridge

IBM literacy support

- 1400 schools in 26 countries
- $5 million grant investment
- IBM annual grants for technology setup
- Includes 85 children's books, 170 adult books, & ELT stories

"Reading Companion has opened new cultural horizons for our children. With such a wide choice of books to increase their vocabulary and improve their comprehension skills. They're developing a true love for reading."
Patricia Díaz Covarrubias, Executive Director, Christel House de México, A.C.

CAMBRIDGE ENGLISH
Language Assessment
Part of the University of Cambridge

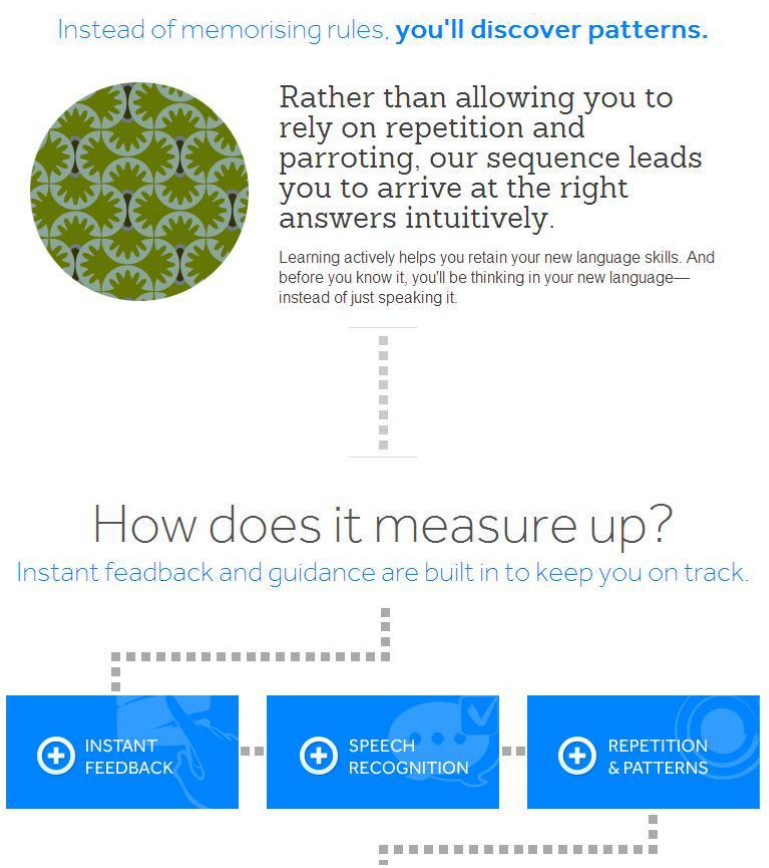
4-How can we use it in c

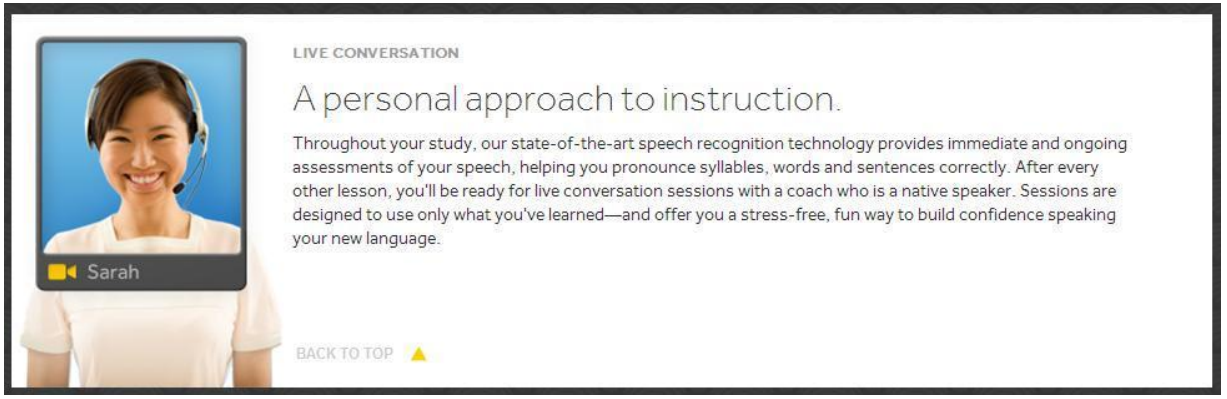
Teach

Learn

Assess

Write





Teaching

Pronunciation

Early spectograph comparisons - inaccurate

Feedback loop

CAMBRIDGE ENGLISH
Language Assessment
Part of the University of Cambridge

Learning

Phonology

Reading companion

CAMBRIDGE ENGLISH
Language Assessment
Part of the University of Cambridge

Writing

Dragon Dictate

ASR in the classroom

If students have Siri or similar:

They tell a story by dictating to machine

One student as dictating role?

Group edits the resulting text and checks accuracy

CAMBRIDGE ENGLISH
Language Assessment
Part of the University of Cambridge

ASR activity

SS write a dialogue

Perform it as dictation

Correct written output

Open conversation

Take in turns to dictate response to previous student

CAMBRIDGE ENGLISH
Language Assessment
Part of the University of Cambridge

ASR self-study

Tr gives text or dialogue to practice

St practises dictating it – checking output measures the teacher model (listening to comparative audio if available)

Futuristic ASR (next year)

Ss have open conversation/dialogue and ASR converts to text, lets them repeat if they are not happy, then emails text of speech to teacher, along with audio of conversation – teacher can grade text quicker but can sample audio

Ss respond to speech prompts with new speech, which ASR converts & translates back to L1 for checking

Weaker Ss speak in L1, hear, L2 translated in ear, repeat L2 and see it ASRed for checking

Ss speak L1 to watch/glass/earpiece and hear L2 in ear, for repeating and internalising;

CAMBRIDGE ENGLISH
Language Assessment
Part of the University of Cambridge

Reflection

How would you use ASR in your class?

What would you need to make it helpful?

CAMBRIDGE ENGLISH
Language Assessment
Part of the University of Cambridge

5-ASR and Sp2Sp translation

Google Translate app

Phrasalator

Rosetta Stone

Google Glass

CAMBRIDGE ENGLISH
Language Assessment
Part of the University of Cambridge

How Google Translate works

- When Google Translate generates a translation, it looks for patterns in hundreds of millions of documents to help decide on the best translation for you.
- By detecting patterns in documents that have already been translated by human translators, Google Translate can make intelligent guesses as to what an appropriate translation should be.
- This process of seeking patterns in large amounts of text is called "statistical machine translation". Since the translations are generated by machines, not all translation will be perfect.
- The more human-translated documents that Google Translate can analyse in a specific language, the better the translation quality will be. This is why translation accuracy will sometimes vary across languages.

CAMBRIDGE ENGLISH
Language Assessment
Part of the University of Cambridge

How Google Translate works….

Google Translate's M.O. consists of sifting through large piles of data — in this case, text. Google refers to this process of translation by finding patterns in vast swathes of writing "statistical machine translation."

As humans, when we learn languages, we do so by navigating the sets of rules which govern them, so Google's process might seem deeply unintuitive.

However, when you compare its results to those of translation services like Babel Fish, which is powered by the rule-based machine translation of SYSTRAN, the improved accuracy of the results speaks for itself. Indeed, Google used SYSTRAN for its translations up until 2007, when it switched to its own system.

At the time, Google research scientist Franz Och explained the switch as follows:

"Most state-of-the-art commercial machine translation systems in use today have been developed using a rules-based approach and require a lot of work by linguists to define vocabularies and grammars. Several research systems, including ours, take a different approach: we feed the computer with billions of words of text, both monolingual text in the target language, and aligned text consisting of examples of human translations between the languages. We then apply statistical learning techniques to build a translation model."

CAMBRIDGE ENGLISH
Language Assessment
Part of the University of Cambridge

Reflection

What does the instant availability of on-demand speech-to-speech translation mean for your teaching and your students' learning?

CAMBRIDGE ENGLISH
Language Assessment
Part of the University of Cambridge

6-Using Sp2SpT in class

It is happening so we need to make space for it in our approach

GRAPHIC

Learn-perform orally-check meaning via S2S translation – discuss differences in group/with teacher

n.Fluent
Translation to the power of n.

CAMBRIDGE ENGLISH
Language Assessment
Part of the University of Cambridge

Using Google Translate

SS write a sentence or short text in L1

StA translates it into English

StB speaks it into Google Translate

Students compare the outputs and note differences, asking for teacher guidance where needed

CAMBRIDGE ENGLISH
Language Assessment
Part of the University of Cambridge

7-ASR auto-marking of speech

How does it work?

ASR conversion to text

Process and analyse

Language model?

Compare to corpus?

Carnegie Speech

Students study in class and practice at home – where they speak into the microphone and get feedback on pronunciation, stress & intonation performance

Claims to understand word meanings, but patchy

CAMBRIDGE ENGLISH
Language Assessment
Part of the University of Cambridge

Automatic grading projects

iLEXir

- have developed an automated ESOL text grading system, to which speech grading is being added

CANTAB

-Cantab Research offers large vocabulary speech recognition in British and American English. Working with our customers we have created systems for indexing broadcast speech, the transcription of voicemail messages, medical dictation systems and several novel applications of automatic speech recognition.

-Systems may be created on the customers site or on Cantab's extensive processor farm and either using customer data or drawing on the many large corpora held by Cantab Research.

CAMBRIDGE ENGLISH
Language Assessment
Part of the University of Cambridge

Cambridge ALTA Institute

Cambridge University Institute for Automated Language Teaching and Assessment (ALTA)

A new research institute supported by Cambridge English which will investigate how technology can support language learning and language assessment, in these areas:

- text and speech processing
- machine learning
- corpus development and analysis
- security, platforms and deployment

- Huge advances in areas like 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.

- Automated assessment won't replace human examiners anytime soon, but it can add great value to their work. For example, it can provide additional layers of quality control, speed up processes and allow teachers to offer more objective in-course tests which give detailed diagnostic feedback to help students to improve their English more effectively.

ASR components

Pronunciation Lexicon

Recognition Engine

"The cat sat on …"

Acoustic Model

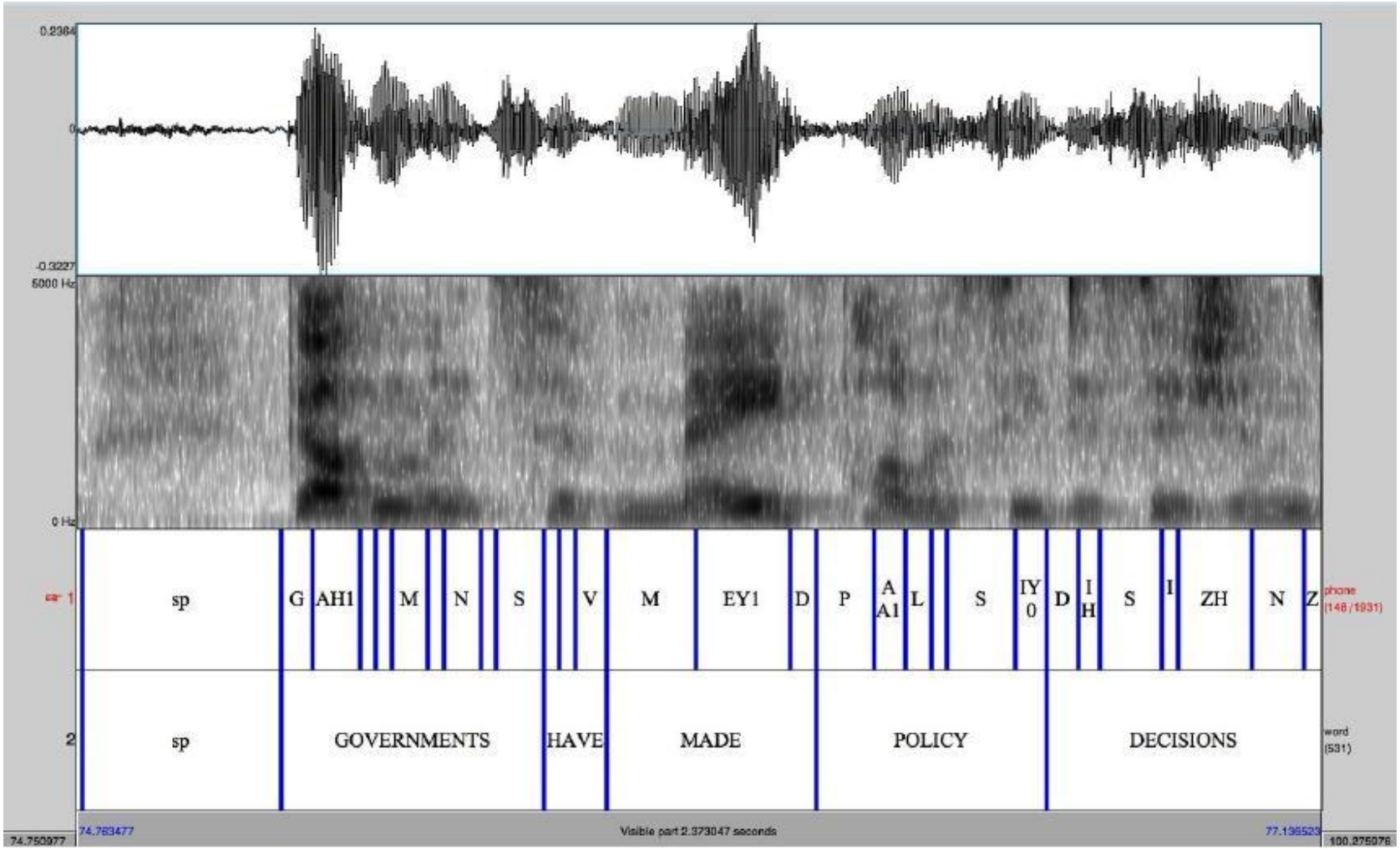
Language Model

Acoustic Model training data

Language Model training data

CAMBRIDGE ENGLISH
Language Assessment
Part of the University of Cambridge

Aligning speech and text



Assessment

Mark Gales video

http://www.policyreview.tv/video/920/6996

ASR Output

yeah actually um i belong to a gym down here gold' s gym and uh i try to exercise five days a week um and now and then i' ll i' ll get it interrupted by work you know

Meta-Data Extraction (MDE) Markup

/{DM yeah actually} {F um} i belong to a gym down here / / gold' s gym / / and {F uh} i try to exercise five days a week {F um} / / and now and then [REP i' ll + i' ll] get it interrupted by work {DM you know } /

Written Text

I belong to a gym down here. Gold's Gym. And I try to exercise five days a week and now and then I'll get it interrupted by work.

CAMBRIDGE ENGLISH
Language Assessment
Part of the University of Cambridge

8-Future trends

Wearables:

Watches

Google Glass

Phone systems

Speech to print output

Speech activated equipment

Widespread auto-marking

Speechprint ID systems

Will ASR replace teachers?

Changing role of teachers?

Shift in status of teachers?

Embracing technology and incorporating it can lead to a higher professional status – in contrast to the t-shirt & jeans image of ELT

Teacher Development Needs?

- Digital literacy development
- Digital pedagogy workshops
- Prepared lesson resources

CAMBRIDGE ENGLISH
Language Assessment
Part of the University of Cambridge

Contacts:

Cambridge English sites:

- www.teachers.cambridgeenglish.org
- www.cambridgeenglishteacher.org

Comments:

Carrier.m@cambridgeenglish.org

If you would like copy of the presentation & references:

www.michaelcarrier.com