LOOK AGAIN

Automatic Speech Recognition (ASR), “to convert speech into a sequence of words by a computer program” (Huang 2010, 339), is back in the limelight [1], thanks to Siri & Co. Artificial intelligence applications like these are impressive, but Second Language Acquisition (SLA) programs will find them difficult to integrate. They could instead reevaluate the underlying ASR, which has undergone two decades of a continuous improvement. This article discusses how ASR can be used in SLA programs, provided the right management. Fortunately, “[s]peech recognition is not about building HAL 9000. (...) Our job is trying to find a good use of an imperfect, often crummy, tool that can sometimes make our life easier” (Pieraccini 2010). The SLA learning application presented here does not pretend to “understand”, and will not be employed to independently grade students. It won't replace, only relieve the teacher from some “traditional classroom-based pronunciation instruction” (Thomson 2011, 748), and especially build “confidence and fluency”. The “simplest [learning] principle is practice makes perfect. (...) A machine that listens is a good practice tool” (Egan 1999, 290).

Have the Learning Resource Center (LRC) Answer SLA Needs

The LRC at a large state school for which I developed this application of ASR is a multimedia-computerized (headphones, digital audio lab software) facility, primarily centered around supporting face-to-face classes, with additional facilities
supporting proctored exams and assignments during self-access. This use of ASR was meant to integrate with other LRC activities for oral practice, assessment and ePortfolio (a lower-key and more frequent homework assignment than our Kaltura student presentation webcam recordings). These assignments prepare for in-class chapter tests with textbook audio recordings, screencast recordings of student presentations, pair conversations, and question-response midterm/final exams using Sanako digital audio lab. All of these also produce artifacts for the institution’s language learner ePortfolio. However, what really reinforced the need for introduction of ASR was the conversion of the 1st-year Spanish program into a “hybrid” mode. Hybrid delivery cut contact hours in half and catapulted the institution’s entire language program to the bottom of the UNC system rankings for average language class size. The online textbook with its speech recording software was to compensate for the drop in contact hours. However, such recordings would have to be evaluated manually. For lack of time, only the self-grading activities were assigned – even though the ADFL keeps admonishing hybrid classes to pay special attention to the development of oral proficiency in order to help compensate for the reduced face-to-face classroom interaction (ADFL 2014). Technology to the rescue? The obvious standout feature of speech recognition is immediate automated feedback via the computer’s transcription “response”. Luckily for high-enrollment programs such as Spanish, Windows 7 Automatic Speech Recognition (W7ASR) was most widely available as a tool to support speech practice activity in the LRC. W7ASR’s only requirements are a PC with Windows 7 and headphones. A W7ASR homework assignment can be flexibly scheduled and was the only way to provide the bulk of our students with guided speaking practice.

Install the Tool

Public awareness of ASR focusses on Siri and a few others. However, mobile phones are unlikely to be mainstreamed in big state schools’ learning environments soon. They also have battery life and bandwidth characteristics that limit continuous ASR. For many institutions, most alternatives are precluded by cost anyway. The free web browser-based Google Translate voice input is, as of this writing, only available in English and many of our teachers dread it because students routinely abuse it as a translation tool instead of using a dictionary tool. Instead, we installed all available language technology add-ons in Windows 7 (and Office 2010) free for Windows Enterprise with volume licensing. Only a few of the language packs available for the Windows 7 MUI include W7ASR (namely Chinese, English, French, German, Japanese and Spanish) since “it takes a global village” of local PC installations to warrant the production cost.

The resulting product consistently ranks next to market-leading (and self-advertising) Dragon Naturally Speaking (1) (Solutions center, 2010).

Additional tools needed:

- MS-Office (or similar, for proofing tools and tracking changes)
- a screencast recorder (free and easy MS-Community Clips)
- an LMS for uploads of one to several dozen MBs files
- the W7ASR Profile Tool (to backup individual voice training data in a shared, “frozen” LRC).

Understand the Tool

While W7ASR was not designed for SLA – and all ASR is still struggling with SLA accents, I found it sufficient for speaking practice in the various cases I tried (unlike the SLA specific Auralog Tell me more (ATMM) ASR, whose recognition accuracy – even though ATMM’s exercise design lowers the bar for ASR – I found very much lacking, see my AATM speech Recognition Test). Another difference to ATMM is that W7ASR does not present the user with an aural model to try and match phonetically. Nor does it allow the user to compare a voice graph of user speech input with a voice graph of a model. However, voice graph visualization is not known to be meaningful for the language learner or to have any benefit for pronunciation improvement (Thomson 2011, 747) – which I could confirm when observing students trying to work with ATMM’s voice graph. With extra work, you could add such an aural model and voice graphs, by using a digital audio lab recorder – either with manual recordings as a model, or even with text-to-speech generated, using Google or Microsoft/Sanako text-to-speech. Still, voice graphs added with the Sanako Student Recorder would be even less useful than ATMM’s. I can only find use in this feature when navigating to beginning of utterances.

W7ASR does not come with learning content either – a blessing in disguise. Despite considerable systems integration work, the ASR of my earlier ATMM installation was hardly used since its content could not be aligned with the existing syllabus and textbooks (and for similar reasons, we faded out a Rosetta Stone installation). Another blessing in disguise is W7ASR’s need for voice training. ATMM had individual user accounts, but its only adaptation to user speech input that I noticed was that ESL learners with especially poor pronunciation, after several non-improving attempts, became suddenly “successful”, or rather began to be ‘waved through’ their speaking exercises. For good recognition results, you have to train the ASR software to your voice (see e.g. Coniam 1998, 20) although Siri & Co. may just hide this from you. Speaker-dependent W7ASR...
sends students through a voice training before the first assignment and after that, additional learning also happens behind the scenes.

Finally, since speech recognition is still hampered greatly by unsuitable acoustic conditions (Baker 2009, Huang 2010), your LRC can help you with carefully controlling the technical aspects that can make such projects fail in language programs with limited resources. Mainly by providing:

- initial language pack installation (or with explaining to IT why an LRC benefits from an install of MS globalized software resources)
- high quality acoustic hardware (directional microphones to minimize background noise, the number of simultaneous users depends on your acoustics)
- audio configuration (right input sensitivity of your microphone)
- user training

Train the Users

For a successful implementation of ASR in your SLA program, it is crucial to provide users with adequate training. For ideas, have a look at my training samples:

- Example of task overview, with a heads-up regarding typical pitfalls:
  - “Activate target window!” (password: wsra7)
  - “Don’t fall back on L2!”, (password:w7ras)
  - “Turn ASR off after exercise completion!” (password: 7awsr)
- Full-cycle hands-on multimedia training materials (samples for various languages [4])
- Example of initial user training:
  - Speech recognition wizard
  - Voice training wizard
  - Backup of Voice profile
- Screencast documenting speaking, transcription and correction while “tracking changes”, to be submitted via a Sanako or LMS file upload assignment or even e-mail (attachment size permitting).
- Dictation and Voice Command tasks

Tip: For the voice training task, non-western L2 learners need about 2 years of prior language study (view Japanese example, password: 7sraw) and western L2 beginners need at least some preparation. Some prompts can be trained for, but also note that the voice training algorithm is adaptive and presents a known user with different prompts whenever the training is restarted.

Integrate Into the Syllabus

W7ASR has 2 modes: Voice command (VC) and Dictation. In VC mode you operate the graphical user interface by voice instead of a mouse. W7ASR’s closed-response design is not only user-extensible (by way of Windows speech macros), but has more choices, some visible (menu names etc.), some hidden and seems superior to ATTM’s (Eskenazi 1999, 464, Thomson 2011, 747), though not designed for SLA.

More suitable for language programs is the Dictation mode which replaces keyboard with voice and is a large-vocabulary or “open-response design” (Egan 1999, 287). Using Dictation resource kit, you can even open this up further with your own domain-specific vocabulary. Dictation mode makes W7ASR also a ‘continuous ASR system’. Microsoft advised to speak naturally and fluently and not over-enunciate – I had best results when I upped my enunciation a bit. W7ASR can return to “isolated word” (ibd., 253) or “discrete word ASR” (ibd., 255) – not trivial, given the inner probabilistic workings, but useful during corrective passes over misrecognized words (see sample here @“User corrects with speech”, pwd: w7ras).

Most important for us, right behind the liberal licensing, and yet another way to minimize total cost of ownership, is the easy integration of W7ASR with the textbook-based syllabus (at least until online textbooks add – after audio players, then recorders – ASR software). Here are some suggested uses:

- My initial test of W7ASR applied it to an end-of-chapter webquest from a 1st-year German textbook. From reading for the gist to pronouncing an authentic target culture text is a challenge – make it extra credit, since reading aloud is being reappraised as an L2 learning strategy (Gibson 2008).
- Turn existing writing exercises into speaking, reading, corrective writing exercises:
  - Free form writing during note taking, e.g. when answering comprehension or short essay questions, fits the...
continuous speech optimization of W7ASR.

- You will likely see reduced recognition accuracy and less proofing feedback, but gain wider applicability for grammar, but especially vocabulary drills that use discrete writing, e.g. filling in cloze exercises (navigation need not be done by voice), even in a web browser from your LMS or online textbook.
- Turn “flipped classroom” homework and fill-in-the-blank conversation suggestions into more instructional phrase dictations to prepare students for your in-class communicative practice. Most 1st- and 2nd-year textbooks prompt students to work in pairs to form small question-response dialogues using building blocks, thus practicing structures and vocabulary while gaining conversational fluency. Instead of only filling in the blanks in writing during homework, using ASR, students can also practice pronunciation at home, and bring their results (be it in print or as MS-Word web app) to review during class, focusing on the conversational aspect there.
- Similarly, my Spanish screencasts use an example from our Spanish for Law Enforcement course: If the computer can ‘understand’ you reading the Miranda warning, a citizen likely can as well.
- Finally, the most advanced language learners can replace writing of essays by speaking them – use of only ASR for editing optional.

Our task design tries to be multifaceted, multimodal and reinforcing, by combining some, possibly all of the 4 skills, listening (to a model or the own pronunciation), (re-)reading (source text and recognized text), speaking, and some writing (unless student prefers to correct with speech recognition). W7ASR can not only grow on you, but it can grow with you and accompany learners throughout their studies. Students are encouraged to update their speech profile (on thumb drive or cloud) after each session, since W7ASR continuously learns to improve from user speech. Students can collect their best assignment submissions for the language learner ePortfolio – a potential employer cannot judge your accuracy, but your screencasts show that you make a machine type a letter for you in L2 with no proofing errors. At the other end, even a 1st-year, possibly 1st-time language learner can experience, while working in a private, low anxiety environment, how talking (using voice command or simple dictation mode) can control a machine, making it do things with “incantations”.

Grade the Assignments

For grading, we provide the teacher with both maximum simplicity and, if desired, additional evidence of their students’ learning. All speaking assignments are documented through uploaded screencasts of the speech, its transcription and manual correction. The teacher could give credit for the following:

- just the submission of a screencast, honoring the learner's effort practicing.
- the number of written corrections the learner needed to make (easily visible, thanks to “track changes”) in the last frame of the screencast, or, if questioning ASR validity, rewind the screencast to examine more closely a misrecognized word.

To provide additional help through LMS assignment comments, recheck the entire screencast – student can only benefit from the human refinement, after already benefiting from immediate automatic feedback. However, there is no need to closely assess actual pronunciation accuracy, if we accept W7ASR's language model as a cost-free approximation of what constitutes “comprehensible output”, for it is speaking “practice [that] makes perfect”.

GO FORTH

We just upgraded to Windows 7, and, while my first demo of W7ASR on an LRC computer dates back to summer 2012, I held my first faculty workshop on this in spring 2014. I am looking forward to learning about teachers’ innovations in the improved Windows 8 setup (I recently installed the Windows 8 German Language pack on my Windows 8.1 PC in less time than it took me to blog it). This update could take the application of W8ASR as a “machine that listens” in language programs outside of the confines of the LRC and onto your students' very own PCs.

REFERENCES


[1] As exemplified by Gartner’s current “hype cycle” (Babbage 2014), according to which ASR approaches the “plateau of productivity”.

[2] Among which I count myself, having been tasked with hosting an ATMM Tell-me-more (ATMM) in the language center of a UK School in 2006.

[3] See e.g. how Brian McFadden seems to assume everybody can sympathize with John F. Kerry when his “IRAQ WAR 3.0” strip for nytimes.com from September 14, 2014 climaxes in Siri, billed as “Enhanced Military Intelligence”, responds to the Secretary of State's inquiry “Can't we solve this diplomatically?” with “I am unfamiliar with that term. Did you mean ‘Airstrikes?’” (direct link beyond the strip's home page seems not possible).

[4] You are also welcome to reuse my entire slide deck for my speech recognition training workshop (under CC BY-SA 2.5).
Speech recognition tool which provides translation of text into audible voice recordings through automation. Learn more about SESTEK Speech Recognition. Efficient note-taking: A common misconception around speech recognition solutions is that such tools are error-prone. However, as speech recognition systems approach near-human levels of accuracy, this concern has become virtually nonexistent. In fact, users now look at these solutions as a way to improve accuracy in their note-taking and documentation processes. Free online speech recognition tool that will help you write text with your voice without typing. Speech to text notepad. Transcribe audio and video files. Speech Recognition Tool. This utility is a speech/voice to text reader. What does it mean? It means that you can just dictate a text and it will be written without lifting a finger. It has the following features: Free and online. No downloads, installation or registration. Cross-platform. Multi-language support. You can pause or stop dictation (it saves the position of the last word).

Try the Yabla Player

Spanish  French  
German  Italian  
Chinese  English  

yabla