SPEECH TECHNOLOGY FOR LANGUAGE LEARNING: RESEARCH & TODAY’S TOOLS

OVERVIEW

Due to rapid advances in speech technology in recent years, tools for language learning are also transforming. This is seen through speech technologies such as Siri, Alexa, and Google Assistant (Apple, 2016; Amazon, 2014; Google, 2012) which include automatic speech recognition and speech synthesis with text-to-speech. For language learners, research demonstrates the efficacy of technology for second language (L2) speech development and pronunciation growth (Bajorek, 2017a). Effective studies include visualizations, feedback to learners, and scaffolded activities that build awareness of linguistic and phonetic knowledge.

In this article, contemporary research and tools are explored and connected to practical teaching techniques to try out today. Contemporary free speech technology tools like ELSA and Google Docs Voice Typing, which can be used in the classroom, provide innovative diagnostic tools and voice typing available in 119 languages (Ong, 2017). This software expands learner insight into speech patterns and overall support of language exploration, leading to language acquisition.

TOPICS AND QUESTIONS COVERED:

1. Speech Technology
2. Best Practices Recommendations Checklist
3. Speech Development and Speech Technology
4. Today’s Tools
5. Pronunciation Goals for Learners
6. Future Directions

WHAT IS SPEECH TECHNOLOGY? SEEMINGLY “MAGICAL” TECHNOLOGY

In the past few years, contemporary speech technology has been radically evolving as seen through commercial products such as Siri (Apple, 2016), Alexa (Amazon, 2014), and Google Home (Google, 2012). What was once rejected as awkward and “nearly unusable,” speech recognition is now used for automating tasks such as writing emails and dimming the lights (Johnson, 2013, para. 3). There is growing ubiquity of these products. The growing presence of speech technology is seen in how they are being bought and brought into homes more and more. It is estimated that one in six Americans owned a “smart speaker” after the 2017 winter holiday season (NPR, 2017) and these products are expanding globally. This is a 128% increase from January 2017 (NPR, 2017). Due to rapid improvements related to machine learning, big data, and computational innovations, this software can seem almost “magical” in nature (Johnson, 2013), but this is only first few pages in the story of speech technology.

Figure 1: Smart Speaker Ownership (NPR, 2017, Slide 5)

These swift advances in technology are directly connected to the ability to speak clearly and effectively, whether in your mother tongue or fourth language. Incidentally, intelligible speech is of immense importance to second language (L2) learners, though it is not always supported in the modern L2 classroom (Bajorek, 2017a). Moreover, the development of L2 pronunciation, speaking and listening skills, is crucial for overall acquisition of spoken languages. Today, advances in L2 pronunciation research explain the state of the field and how L2 instructors and learners can be supported by speech technology.
Speech technology is the processing of speech by computers, divided into the two subdivisions, speech synthesis and speech recognition (Levis & Suvorov, 2013; Rodman, 1999). Speech synthesis generates intelligible human speech from text, while speech recognition is software that interprets human speech sounds acoustically and creates text through natural language processing, machine learning, and other computations. While both are forms of speech technology, speech recognition has more utility to language learners who need to practice speaking themselves, though speech synthesis (also known as “text-to-speech” or TTS) has been found to be effective for learner development of speech also (Liakin et al, 2017). Specifically, automatic speech recognition (ASR) gives speakers almost instantaneous feedback about what the computer did or did not understand of their speech.

**TESTIMONIALS:**

As speech technology is incorporated into language learning environments, consider the below as guiding frameworks for how best to talk about pronunciation and L2 speech development with L2 learners.

**Focus on Learner-Centered Goals:** What do learners value? What skills will they need in the real-world? Does instruction support the goals of learners?

**Intelligible vs. Nativeness:** Be open to dialogues about why speaking clearly and being understood is more important than sounding like a native speaker.

**Identity, Pronunciation, and Dialogues:** Learners may have strong reactions that are related to their identity when talking about these subjects of dialects, accents, and pronunciation. Be thoughtful about when and how to have clear and supportive dialogues about speech and pronunciation.

**Acceptance of Pronunciation:** The topic has been long-neglected (Morin, 2007). Instructors may feel unprepared to teach this topic, but should consider trying out new technology that can support learner speech development.

**Test out Tools:** Strongly recommended that instructors test out technology before implementing them in the classroom. Tools vary significantly and may be a perfect fit or not quite the right tool depending on learning contexts.

**Set Reasonable Expectations:** Technology can seem attractive just because of aesthetics and hype EdTech. Technology itself is not a silver bullet and instructors and students should recognize both its limitations and its benefits.

**Advances in the Future:** In the next 5-10 years, speech technology tools will grow and change rapidly. Expect updates in the future about education-specific software to support L2 speech development.

**SPEECH DEVELOPMENT & SPEECH TECHNOLOGY**

Current research indicates that speech technology can absolutely support L2 speaking and listening skill development through feedback (Golonka et al., 2014; Liakin, Cardoso, & Liakina, 2015; Lee et al., 2015; Thomson & Derwing, 2014). Many instructors are aware that speech technology could be helpful, but the remaining question is about how to use it effectively. There are so many new products arriving daily and there is such little evidence about whether these products can support learning. Overall, contemporary research indicates that speech technology that gives feedback that is specific, evidence-based, and actionable, can dramatically improve L2 learner pronunciation (Bajorek, 2017a). When in question, look for products that include visualizations, feedback, and data-driven tools.

For pronunciation, a common speech issue with learners of English is the famous vowel confusion with “sheep vs. ship.” While some may be aware that they struggle with some vowels in English, it is completely different from having a vague idea and seeing their speech diagramed by technology. Through visuals, learners can work to move their speech towards meaningful pronunciation goals (Carey, 2004; Kartushina, 2015). Below are figures of effective techniques used in research, however these tools were designed for small-scale studies and are not available to the general public. In Section 6: Today’s Tools, there is a discussion of tools available for language instructors today.
Figure 3: Text-to-speech Exercise Example, Reproduced from Liakin et al. 2017, p. 355

Using text-to-speech software, French learners who used carefully designed activities for noticing words and pronunciation features improved their production of French liaison (i.e., je suis ici) in around 2 hours (see Figure 3) (Liakin et al. 2017). These activities were designed with the commercial software NaturalReader (2018).

While these are only two examples of the many speech technology tools available for language learning development, they illustrate ways in which these tools can support speaking development. With any tools, it is important to specify the goals of using speech technology in L2 development.

**TODAY’S TOOLS**

Understanding the research at the interface of speech technology, SLA, speech and pronunciation is important when considering the application of speech technology to language acquisition. As mentioned in Section 3, although not all tools that are found effective in scholarly research are available to instructors, there are some speech technology tools for language learning that are available today.

A word of caution however, not all of these tools have been researched and many rely on customer reviews rather than empirical evidence that demonstrates their efficacy.

Prominent language learning software cannot always be trusted in this domain. In a review of Rosetta Stone, Duolingo, Babbel, and Mango Languages (Bajorek, 2017), the use of speech technology in these tools was found to be largely unhelpful for language learning.

This is significantly problematic since hundreds of millions of people worldwide use this type of software, for which some are willing to pay. While Babbel has sophisticated automatic speech recognition software, it is strongly recommended that other tools that are specifically focused on supporting language learner speaking development are prioritized over these more commercial options (Bajorek, 2017).

To aid this, this section specifically looks at free and state-of-the-art speech technology: ELSA and Google Voice Typing. Possibilities for use of these tools for education abound. As tools are designed, specific goals should be designed with objectives of the learner in mind.

ELSA (English Language Speech Assistant) which is the most sophisticated pronunciation feedback system on the market (Bajorek, 2017a; Van, 2018). Although competitors are now filing in, i.e. Blue Canoe Learning (Daniels & Andrews, 2017) and SpeechAce (Gupta, 2018).

ELSA is a paid service and its functions are currently only available for English as a Second Language (ESL) learners, but there is great potential for other languages to be supported in the future. ELSA was designed to give learners specific feedback about their speaking using state-of-the-art artificial intelligence automatic speech recognition (ASR), seen in Figure 4.

Targeting specific sounds, the app analyzes which speech sounds you make that are not like a “native American English speaker.” No research has been found regarding the efficacy of the software or the accuracy of its assessment measures, but the company’s website is rich in customer reviews (Van, 2018).
In Figure 4, several aspects of the app are presented. Color-coded phonemes in different contexts in a word are provided through extended practice with the app (Figures 4A and 4B). International phonetic alphabet (IPA) transcription is available (Figure 4B), but may not be understandable to all users. Overall, ELSA relies heavily on IPA letters and transcriptions as well as percentage accuracy rates, seen in Figure 4C. While there are no explanations for how the percentages are assigned, these may be encouraging or discouraging to users depending on how they feel about their “American accent.”

**GOOGLE DOCS: VOICE TYPING**

Google Voice typing is backed by the most advanced software in the field for automatic speech recognition. Considered one of the best in the world, Google reported a 92% word accuracy rate for its speech recognition technology in 2015 (Novet, 2015; Tatman, 2017). A lesser-known and free service provided by Google is the voice typing capabilities built-in to Google Docs. As seen in Figure 5, this service can be helpful for language learners as they see the direct connection between pronunciation and writing. Google recognizes 119 language varieties and claims that using voice typing is three times faster than regular typing (Ong, 2017). The software gives learners direct feedback about what the automatic speech recognition can and cannot understand. The service is free, but this does mean that learners are giving away their personal data.

**PROS:**
- Free and easy to use
- Supports 119 languages and many dialects (i.e., 22 varieties of Spanish)
- Immediate feedback

**CONS:**
- Users must set up Google Accounts
- Software is not perfect: The software is not perfect and may not understand all speech, even for native speakers of a specific dialect. This topic should be considered seriously before using this tool with language learners.
- Software Requirements: This technology may only run on certain computers that have enough working memory and permissions to use the microphone.
PRONUNCIATION GOALS FOR LEARNERS: OPPORTUNITIES FOR SPEECH TECHNOLOGY

On worthy area of focus of speech technology is pronunciation. There are many reasons why pronunciation instruction and development is neglected in the typical L2 classroom. Support for L2 pronunciation in classrooms is frequently overlooked (Trofimovich & Isaacs, 2016) and terms such as “neglected” (Elliott, 1995, p. 530), “casualty” (Thomson & Derwing, 2014, p. 326), and “stepchild” (Arteaga, 2000, p. 340) have been used to describe the state of the field today.

This can happen for a number of reasons, for example, instructors may not have a linguistic background or training in pronunciation (Morin, 2007) or students may feel anxiety around speaking in their L2 (McCrocklin, 2016). However, the single most attributed reason to the neglect of L2 pronunciation in the classroom is related to the theory behind contemporary SLA classrooms (Bajorek, 2017a; Thomson and Derwing, 2014). In the evolution of the SLA field, Krashen argued that L2 pronunciation would develop naturally over time and implicit understandings would arise with understandable L2 input; there was no need to explicitly address pronunciation within the language classroom (Jones, 1997; Krashen, 1982; Thomson & Derwing, 2014). However, this idea is directly contradictory to a large body of literature indicating that listening to the second language alone is insufficient for pronunciation development with research measuring potential changes over long timeframes, spanning from 12 weeks to 4 years (Elliott, 1995; Elsendoorn, 1980; Flege, 1981; Flege & Hammond, 1980; Han & Odlin, 2006; Mitleb, 1981; Niemi, 1979; Solon, 2016; Waniek-Klimczak, 2013).

This mindset of “listening is enough,” led to the “virtual disappearance of pronunciation work” from textbooks of the 1970’s (Jones, 1997, p. 105). Since then, relatively little has changed in mainstream L2 pedagogical material regarding pronunciation (Arteaga, 2000; Lord & Fionda, 2013; Terrell, 1989). Prominent L2 educational technology (EdTech) tools do not support L2 pronunciation development either, as seen in a 2017 review of Rosetta Stone, Duolingo, Babbel, and Mango Languages (Bajorek, 2017a). Fortunately, a shift is underway to focus more on pronunciation and supporting pronunciation learning in academia and commercial products.

With this renewed focus, we must remind ourselves, that all language education should focus on the goals of learners. If the learners would like to use language in the real world, practicing words and sounds that are used frequently which are meaningful for contrast in speech, these are the features that should be prioritized. This leaves us with a reiteration of the important question, which “problems would you tackle first?” (Brown, 1974, p. 53 cited in (Brown, 1988, p. 593)).

In L2 pronunciation research, the concept of “functional load” ranks which speech segments are more important than others in a specific language (Brown, 1988, 1991; Catford, 1987). Errors that are particularly salient and common are the ones that instructors might want to design lessons around before moving on to ones that are less frequent (Eskenazi, 2009). Most research about prioritization and functional load in pronunciation has focused on English and the importance of consonant contrasts in English (Celce-Murcia et al., 2010; Kang & Moran, 2014; Sewell, 2017). Clearly, this prioritization of concepts also applies to all languages.

Research in this subsection of the field is scarce. However, what is available, on relative functional loads for English vowels (Celce-Murcia et al., 2010, p. 160), for example, tells us that /I/ and /æ/, as in bit and bat, have a relative load of 100%, rendering such vowels quite important. Others such as the difference between “bead” vs. “beard” are less important for being understood in English. Until more research is conducted in this section of the field, the idea of how common sounds are should be a consideration of instructors and pedagogical designers.

FUTURE DIRECTIONS AND CONCLUSIONS

Contemporary speech technology has been radically evolving as seen through commercial products and the evolution of this software is only beginning. As indicated by commercial products, smart speakers and speech technology will become more and more ubiquitous in everyday life. Automatic speech recognition and text-to-speech capabilities are more and more ubiquitous and advanced in their capabilities (NPR, 2017). Research demonstrates how speech technology can support L2 speaking and listening skill development through feedback to users (Golonka et al., 2014; Liakin, Cardoso, & Liakina, 2015; Lee et al., 2015; Thomson & Derwing, 2014).

Available today, speech technology ELSA and Google Docs Voice Typing are free and fantastic options for language instructors and learners to explore L2 pronunciation and their voices. Empowering language learners to be able to...
speak more clearly and improve speech features should be the focus of language learning education. Beyond hype around technology, it is crucial to keep a clear-sighted mindset of what is best for learners and learner-centered goals. Speech technology is a tool, not a silver bullet "no matter how sophisticated or effective" the software is (Chun, 2012, p. 10; Jensen, 2017). If well-designed, speech technology can be tailored to specific goals and to enhance existing lesson plans and curriculum (Blake, 2013).

It is time to harness opportunities and insights from research to bring practical and affordable tools to instructors and language learners everywhere. Keep an eye out especially for software that provides clear visualizations, feedback to learners, and scaffolded activities that build awareness of linguistic and phonetic knowledge. It is an exciting time in this field that has such promise for the field of language education in years to come.

SUGGESTED FURTHER READING LIST

Review of L2 Pronunciation in Prominent Language EdTech Software

Overview of Automatic Speech Recognition for Applied Linguists
REFERENCES


Writeup https://www.nationalpublicmedia.com/smart-audio-report/


Joan Palmiter Bajorek

Bajorek is a Speech Technologist and PhD Candidate at the University of Arizona. Exploring virtual reality technology and speech recognition, she is the Principal Investigator of an international educational technology research team in collaboration with ImmerseMe. She is the 2018 Winner of the Outstanding Graduate Student Award from the international organization Computer-Assisted Language Consortium (CALICO). Presenting and publishing at national and international venues, Bajorek’s work explores EdTech, user experiences with technology, and data-driven education. She holds a MA in Linguistics from the University of California, Davis.