MEASURING NATIVENESS FOR THE ENGLISH LANGUAGE LEARNER: EMERGENT TECHNOLOGY FOR THE CLASSROOM

Kenneth Spiegel, MA, CEO[1], Susan Homan, PhD[2]

[1]*LanguaMetrics, Inc. (UNITED STATES)*
[2]*University of South Florida (UNITED STATES)*

ABSTRACT

Speech recognition technology is all around us today. We see it in our phones, our cars, and our living rooms. Why isn't it helping English language learners in the classroom? This paper will explore the reasons for this and how early stage products are gaining ground in China and may soon be helping learners worldwide. We will begin by understanding the difference between speech recognition and speech scoring and why speech scoring is what is needed for English language learners. SRI International (formerly Stanford Research Institute International) of Menlo Park, California is considered the pioneer of speech scoring technology. LanguaMetrics, Incorporated (Inc.). of Tampa, Florida licenses speech scoring technology from SRI International and has created a learning platform for publishers and institutions to create products to help English language learners improve their ability to communicate. The platform empowers its products to actually measure the nativeness of the speaker's oral production. This nativeness measure encompasses the pronunciation, rhythm, stress and all factors that impact the speaker's ability to be understood by a native speaker. Visual feedback is provided to the learner so that for the first time learners can actually see how their speaking can be improved. Implications for the classroom will be discussed with input from Dr. Susan Homan professor emerita University of South Florida.

Keywords: Speech recognition, language technology, language learning, nativeness, English language learners, SRI International, LanguaMetrics, Inc.

1 INTRODUCTION

Over the last decade, speech recognition has grown in use and is apparent in everyday activities, from calling utility providers and using voice commands to operate hand-free devices, to using new smart technology, such as Amazon Echo®. Furthermore, governments around the world have incorporated speech and voice technology which allows them to listen to outside conversations and search for keywords within those conversations.

According to Tractica, a marketing intelligence firm based in the United States (U.S.), the use of voice and speech technology has grown tremendously worldwide, with the Pacific regions being the projected highest growth areas [1]. Additionally, Tractica estimates that the economic growth of speech and voice recognition will increase from $249 million United States dollars (USD) to over $5.1 billion USD by 2024 [2].

Although there is exponential growth in voice and speech recognition technology, this field has yet to be widely used in the classroom, specifically with English language learners (ELL). Education is one of the last areas to benefit from advances in technology, due to several factors: 1) educational funding being tied to government bureaucracy and inertia, causing slow progression, 2) governmental regulations inhibiting education from receiving adequate funding sources, 3) technology and presumed risk being averse in order to protect children, which leads to no incentives to innovate, and 4) large technology companies and vendors with expansive lobbying budgets dominating the implementation of technology, which oversaturates the technology fields, pushing out smaller start-up companies.

However, because of the need for speech recognition for ELLs, China has been one of the early adopters of this technology in the classroom. It is anticipated that, with China embracing this technology, other countries will become aware of the need and capacious benefits to ELLs.

2   HISTORY AND IMPORTANCE

2.1   Distinguishing speech recognition and speech scoring

Speech recognition software takes information and commands from the human voice and returns the requested information and commands. In order to respond to the voice commands, the first step that takes place is the conversion of the human voice signal into plaintext. The plaintext is then processed accordingly to drive commands from the device, such as searching for a phone number, finding out the weather for the day, or texting someone from the contact list. In broad terms, this first step is referred to as speech-to-text [3].

This paper addresses a different area related to speech-to-text; a more specific science called *speech scoring*. The field of speech scoring is dedicated to a very different purpose, as it has different applications in learning to speak and read another language. The field of speech scoring was pioneered by Stanford Research Institute (SRI) International of Menlo Park, California [4]. The U.S. Department of Defense was in need of computer-based technology for the purpose of measuring spoken English and comprehensibility by allied soldiers from non-English speaking countries. The Defense Advanced Research Projects Agency (DARPA) provided funding to SRI International to develop a program that could gauge spoken language and the level in which the soldiers could be understood. After several years of research and development by SRI International's world-class linguists and computer scientists, the science of speech scoring was attained.

The key differentiator is the science of speech scoring – the objective is to measure the understandability of the speaker to a native listener. SRI International designed technology and acoustic models used to return numeric scores on the "nativeness" of speech. Nativeness is the ability of a non-Native language speaker to be understood by a Native listener. The science of speech scoring measures nativeness and applies quantitative analysis, or scoring, in order to measure it. This scoring allows the language learner to make improvements to the language they are learning in order to be better understood in the real world.

The text from language learning is stored in a format called a *grammar*. A grammar, in the realm of speech scoring, is linked to an acoustic model for the particular language, as well as other technology, so that it can literally score the speech. It creates a resulting file of data that contains scores for various elements often including sentence, word and phoneme level scores [5]. An example of such a data set is shown in Fig. 1.

word	word score	word posterior	word confidence	phonemes
the	1	8	60	th e dh iy -0.3 0.14
quick	1	-30	18	q u i ck k w ih k -1.33 0 -0.52 -1.78
brown	2	16	70	b r ow n b r aw n -0.03 0.37 -0.04 -0.15
fox	1	-49	8	f o x f aa ks -2.89 -1.78 -0.16
jumped	1	-14	32	j u m p e d jh ah m p - t -0.92 -0.26 -0.67 -0.57 1 -0.33
over	1	2	53	o v er ow v er 0.22 -0.62 0.13
the	1	3	54	th e dh iy -0.37 0.03
fence	2	17	71	f e n c e f eh n s - -0.32 0.46 -0.13 -0.07 1

Fig. 1: Speech scoring dataset

Score	Range	Comment
Sentence Confidence	1 – 100 (in percent)	Would this sentence be understood?
Sentence Score	1 – 6 (3 = Native, 6 = Perfect)	3 and above will be understood by native listeners.
Word Score	1 – 6 (3 = Native, 6 = Perfect)	3 and above will be understood by native listeners.
Word Posterior	-30 to +30 (negative score is excellent indication of wrong word uttered)	Machine scores for computational purposes
Word Confidence	1 – 100 (in percent	Degree to which word uttered is the word expected.
Phonemes	-1.5 to +1.5	Pronunciation measure

Fig. 2: Decoding a speech scoring dataset.

2.2 Need for speech scoring technology

The number of ESL students in North American colleges and universities has doubled in the last thirty years [6]. Non-English speaking countries struggle to prepare their students to communicate efficiently when studying abroad or accepting jobs in English-speaking countries. Students are required to pass English proficiency tests, such as the Test of English as a Foreign Language (TOEFL) and the International English Language Testing System (IELTS) for the purpose of entering universities in English-speaking countries [7]. Overall, preparation strategies have been developed so that students can achieve good scores on these exams [8, 9]. Students who have done well on these standardized English tests are often admitted to universities in English-speaking countries, yet have difficulty understanding their instructors and speaking in class and social settings. These basic interpersonal communication skills (BICS) are often overlooked or underdeveloped during the ESL learning in the respective native country [10-12]. These students have high academic knowledge - as evidenced by their standardized test scores – but struggle to speak the language in a real world setting, as evidenced by the unfortunately high percentage of foreign students who do not complete their studies abroad [13-15]. An article in a Chinese online newspaper, *People's Daily Overseas Edition*, speaks to the struggles that Chinese students have while studying in the US. The students reported they had difficulty in class due to the speaking speed of the professor and the various English accents encountered [16].

Liu conducted survey-based research on the English competence of Chinese students who prepare for graduate school in a foreign country where English is the native language. The research focused on four areas: 1) self-assessment on English listening and speaking competence; 2) assessment on cognition of English listening and speaking competence and learning psychology; 3) assessment on English listening and speaking text materials and pedagogy in China. The students' responses reveal: 1) 100 percent of them think their listening and speaking competence is not good and believe that they have some communicative obstacles, 2) in China (versus its foreign counterparts), the class size is too big due to fewer classe being offered, there appears to be a lack of practice opportunities, and multimedia is not widely used, 3) 90 percent of students think that the assessment lacks variety in the method of teaching, listening to, and speaking English (mainly cassette, CD and mp3), and 4) they think that the current

textbooks are poor resources [17]. Other research with Chinese students and ELLs suggests a serious lack in appropriate skills in mastering English on a conversational, real-world level [12, 15, 18].

Furthermore, in another survey completed by a Chinese professor, the students believed that the College English Test (CET) 4 and 6 were just to train them to pass exams. CET 4 and 6 are high-level English tests and a passing grade suggests the students have a high-level proficiency of English [19, 20]. However, as the professor conducting the survey indicated, even those who pass the exam have difficulty using English in a real world setting [21]. It is the ability of using the language that matters, not the vocabulary and grammar for the test [22]. One such measure is the use of multiple choice in language proficiency testing. These tests are designed to "look good" and to "focus on what is considered important in the classroom", as opposed to real-time use of speech [23].

2.2.1   Occupational barriers for adults

The decision to hire someone for a company position may be, to some degree, dependent on their ability to speak a language effectively. In the majority of cases, similar standardized tests are used as a screening mechanism with adults who are ESL learners. [7] One such test, the Speaking Proficient English Assessment Kit (SPEAK), has been used to place ESL learners in appropriate classes and support curriculum improvements. SPEAK, through research and examination of the test, does not realistically measure oral communication competency, leading to invalid results of language fluency and capability to use English in an occupational situation [24].

The European Union (EU) recognizes the importance of a multilingual and diverse work force; proficiency testing programs, aligned to the Common European Framework of Reference, gauge potential employees not only on fluency of a second language, but competency [25]. Many potential employees may arrive for an interview and are unable to communicate at a level that would guarantee them a job in an English-speaking country, although they have scored high on English proficiency tests.

2.2.2   Cultural Barriers

Approximately 18 percent of the U.S. population speaks a language other than English, with 11 percent representing foreign-born individuals. Culture plays a role in how a person learns a language and communicates what they have learned, particularly in a social context [6, 26]. Moreover, lexical and grammatical pieces of language are strongly cultural, leading to a high level of language complexity [6].

Three paradigms can be considered when examining non-native speakers in a US culture: 1) communication apprehension, 2) self-perceived communication competency, and 3) social inclination to communicate. Insufficiencies in oral communication courses for English language learners and appropriate curricula contribute to the lack of oral communication competencies [13, 27]. These apprehensions contribute to the failure of academic success of some foreign students. In a qualitative study among Chinese graduate students, 94 percent of students reported their academic experience was negative and not "enjoyable" due to "tremendous" language barriers [9] [6]. Additionally, study abroad programs were thought to be a "cure all" for language and cultural barriers. However, students often arrive to their host countries linguistically underprepared [28].

It is suggested that improving upon language competencies can lead to greater rates of post-secondary education success in ESL students [6, 9]. Furthermore, improving upon speech can limit biases and preconceptions of native listeners towards the ESL speakers, bettering social interactions [29].

3   APPLICATIONS OF A COMPUTER-ASSISTED SPEECH PROGRAM

Computer applications leveraging speech scoring can help ELL's in two ways: assessment and learning. Learning products that leverage speech scoring can provide the student with visual feedback on their speaking ability and help them become more comprehensible. By building in native speaker example

audio files and combining them with the ability to score speech products, a virtuous cycle of learning is created. A pedagogy can be established that first allows the student to listen to a native speaker model of the reading passage. The student then attempts reading the passage out loud into a microphone. The system scores the speech and repaints the passage by turning the letters that were mispronounced red and the letters that were pronounced well green (Fig. 3). The student can visually assess their mispronunciations and can replay the native speaker model to hear how key words should be pronounced, and then attempt reading the passage again. This model of learning in which the student works to improve speech through visual keys, auditory examples from native speakers, and learning recall encompasses a multi-level learning experience for the student. The LanguaMetrics™' platform is designed to produce products that can conform to this paradigm.

Fig. 3: LanguaMetrics™ LanguaBooks™ example showing highlighted text that was not pronounced correctly.

3.1   Speech scoring technology throughout the world

3.1.1. Uses in China

In China, you see a fairly well-developed educational market in which speech recognition products play an important role. However, veterans of this arena will readily concede that the perception of the marketplace of speech recognition technology is not ready to be used in educational forums for measurement purposes. As an example, young students of the Jiangsu Province are administered a test that includes recording a response to an open ended question such as, "What did you do on your summer vacation?", which presents a steep technical challenge: the system must first convert the recording of the spoken English into text regardless of how well they can speak. The text will need to be scored for

grammar, syntax, and logical structure. Unfortunately, the available technology is not adequate in gauging nativeness of the English – the scores are rather random and students believe that, to do better on this test, they must "speak louder" [30]

Fortunately, with the introduction of the National Spoken English Test (NSET) and a new product line based on LanguaBooks™, the situation has begun to improve. In educational settings, speech scoring and the virtuous cycle pedagogy is based on the aforementioned LanguaBooks™ concept running on the LanguaMetrics™ platform. In China, the product is known as *ReadingMate* and is gaining acceptance in the consumer space and on its way into the public school system.

In development for nearly four years, *ReadingMate* combines the virtuous cycle pedagogy based on the ability to score speech, along with engaging content from Rosen Publishing, a major U.S. publisher. The key factor combining the LanguaMetrics™ technology with content from Rosen is the accuracy of the scoring. The ability to provide the learner with precise feedback about their speaking abilities provides a realistic view of how well they are learning the language. Alternatively, as they work through the course, the technology allows the learner to experience real-time improvement, which has not been available previously.

3.1.2. What has developed in the U.S.?

In 2015, based on the success of the LanguaMetrics™ platform in China, the company formed a partnership in the U.S. with the Northwest Evaluation Association (NWEA). NWEA's flagship assessment product, the Measure of Academic Progress (MAP), is taken by millions of school children in the U.S. every year. [31] Now in partnership with LanguaMetrics™, NWEA is deploying assessment products that measure oral language ability, both in the sphere of reading fluency for all students and English language ability for English language learners. Future plans include the development of learning products based on the LanguaMetrics™ platform.

Lynn University of Boca Raton, Florida has partnered with Kid Orange Technology and LanguaMetrics™ to create the Lynn Early Learning Academy (LELA), which is a virtual preschool for children from four to six years of age. LELA will be available in the fall of 2016. Kid Orange is a licensee of ABC Mouse, a well-known consumer learning product for young children. The LELA offering is unique in that it will include content from ABC Mouse, an assessment from NWEA, and LanguaBooks™ including content from Rosen Publishing. The product is further enhanced by evaluation of students' assessment progress by actual teachers – the teachers can also make learning recommendations directly to the parents via the LELA application.

3.1.3 European markets

LanguaMetrics™ development center is located in Kiev, Ukraine. This has provided the opportunity for partnership with Ukraine in promoting the *Year of English – 2016*, in which LanguaMetrics™ has made a custom set of LanguaBooks™ available to Ukrainian school children. Currently, LanguaMetrics™ is hopeful that, with adoption in the Ukraine, other European countries will follow in applying this technology.

4 CONCLUSION

In conclusion, the goals of this paper are to inform educators and businesses of the value of speech scoring and to provide a sense of its market readiness. It is hoped that educators and business people will be inspired by the possibilities to help children and adults. Whether it is a reader struggling for fluency in his own language or a learner of English who can benefit by visual feedback on his speaking, and well-designed speech scoring technology can be of tremendous value.

Products built on the LanguaMetrics™ platform have the goal of either assessing or accelerating learning. High precision speech scoring leveraging technology under license from SRI International is at the core of the platform. Engaging content from quality publishers increases the engagement level for the students.

LanguaMetrics™ products, after years of growth in China, have entered the U.S. market with one of the most prestigious assessment companies in the U.S. – NWEA. Consumer offerings are under way with Lynn University and other partnerships are being formed. The platform is now being made available to the European market and we invite educators, publishers, and learning companies to partner with us for the benefit of their English language learners.

REFERENCES

[1]   Tractica. (2015). Voice and Speech Recognition. www.tractica.com/research/voice-and-speech-recognition. (May 11, 2016).

[2]   Find Biometrics. (2015). Speech and Voice Recognition Market To Hit $5.1 Billion. http://findbiometrics.com/speech-voice-recognition-market-to-hit-5-1-billion-26111. (May 11, 2016).

[3]   Microsoft Developers Network. (2016). How Speech Recognition Works (Microsoft.Speech). https://msdn.microsoft.com/en-us/library/hh378337(v=office.14).aspx. (May 17, 2016).

[4]   SRI International. (2016). Corporate History. www.sri.com/about/corporate-history>. (May 17, 2016).
[5]   Microsoft Developers Network. (2016). About Speech Recognition Grammar (Microsoft.Speech). (May 17, 2016).

[6]   Huang, J., Dotterweich, E., and Bowers, A. (2012). Intercultural miscommunications: impact on ESOL students and implications for ESOL teachers. Journal of Instructional Psychology, 39(1).

[7]   Roca-Varela, M., and Palacios, I. (2013). How are spoken skills assessed in proficiency tests of general English as a Foreign Language? A preliminary survey. International Journal of English Studies, 13(2), pp. 53-68.

[8]   Vasquez Heilig, J. (2011). Understanding the interaction between high-stakes graduation tests and English learners. Teachers College Record, 113(12), pp. 2633-2669.

[9]   Huang, J., and Rinaldo, V. (2009). Factors affecting Chinese graduate students' cross-cultural learning. International Journal of Applied Educational Studies, 4(1).

[10]   Wang, S. (2012). Language problems of overseas Chinese graduate students and their insights into English teaching for graduate students in China. . Academic Degree and Graduate Education, 4, pp. 35-37.

[11]   Huang, J., and Finn, A. (2009). Academic listening tests for ESOL students: Availability, concerns, and solutions. International Journal of Applied Educational Studies, 6(1).

[12]   Liu, C. (2013). From language learners to language users: A study of Chinese students in the UK. International Journal of Applied Linguistics, 23(2), pp. 123-143.

[13]   Burroughs, N. (2008). Raising the Question #10 Non-Native Speakers of English: What More Can We Do? Communication Education 57(2), pp. 289-295.

[14]   Andringa, S., Olsthoorn, N., van Beuningen, C., Schoonen, R., and Hulstijn, J. (2012). Determinants of success in native and non-native listening comprehension: An individual differences approach. Language Learning, 62(Suppl. 2), pp. 49-78.

[15]   Liu, Y. (2011). Suggestions on how to improve English ability of vocational school students. . Test Weekly, 55, pp. 109.

[16]   Dai, J., and Fangge, W. (2010). How to Overcome the Language Obstacles when Studying Overseas. People's Daily Beijing, People's Republic of China.

[17]   Liu, W. (2012). The research and reflection on the English listening and speaking competence of Chinese students studying abroad. English Teachers, 4, pp. 39-43.

[18]   Liu, J. (1994). English language ability and cultural adaptation of Chinese students studying in the US. Liaoning Normal University Academic Journal 6, pp. 57-59.

[19]   Yu, Z. (2008). Problems in college English test 4 and 6 and the reform paths. Modern Education Science, 1, pp. 93-96.

[20]   Chen, C. (2009). Problems in the English tests in China and suggestions: CET 4 and 6 as examples. . Da Zhong Ke Ji, 1, pp. 168-169.

[21]   Zeng, K., and al., e. (2001). Investigation and thinking on the actual English communicative competence of Chinese students studying in the UK. Foreign Language and Their Teaching, 9, pp. 37-39.

[22]   Limin, P. (2006). Tsinghua University Professor Says CET 4 and 6 should be Terminated. China Daily Youth.

[23]   Fulcher, G. (1999). The 'communicative' legacy in language testing. Systems, 28(2000), 483-497.
[24]   Flores, N. (1992). Multicultural communication: assessing, facilitating, and evaluating compentency-based oral interactions. . Annual Meeting of the Speech Communication AssociationChicago, IL.

[25]   Chetty, D. (2009). Speaking the language. Training Journal, 39.

[26]   Roever, C., and McNamara, T. (2006). Language testing: the social dimension. International Journal of Applied Linguistics, 16(2), pp. 242-258.

[27]   Ozdener, N., and Satar, H. (2008). Computer-mediated communication in foreign language education: use of target language and learner perceptions. Turkish Online Journal of Distance Education, 9(2).

[28]   Kinginger, C. (2011). Enhancing Language Learning in Study Abroad. Annual Review of Applied Linguistics, 31, pp. 58-73.

[29]   Jaber, M. (2011). Native speakers' perception of non-native English speech English Language Teaching, 4(4), pp. 77-87.

[30]   Ding, D. (2012). Personal communication regarding English language learning in China. K. Spiegel, ed., Mai Ruike Education Technology Co., Ltd, Beijing, China.

[31]   NWEA (2016). About NWEA. www.nwea.org/about/. (May 17, 2016).