SOFTWARE REVIEW

VoiceTube

Haeyun Jin, Iowa State University

INTRODUCTION

Previous literature on ESL pedagogy has shown that the traditional practice of teaching English pronunciation to ESL students tended to focus on isolated word- or sentence-level pronunciation (Hsieh, 2013; Pennington & Ellis, 2000) and restricted the teaching to isolated segments rather than the level of suprasegmentals (Lightbown & Spada, 2006). As opposed to such a decontextualized way of teaching pronunciation (Celce-Murcia et al., 2010), other research has emphasized the role of suprasegmentals in enhancing intelligibility (Avery & Ehrlich, 1992), and as areas considered more likely to affect communicative success (Derwing & Rossiter, 2003). Affected by such a redirection of focus, more emphasis was placed on the research and instruction on teaching intonation, stress, rhythm, and pause. Empirical research has noted that suprasegmental features, especially intonation should be taught at the discourse level (Chun, 2002; Jenkins, 2004; Levis & Pickering, 2004) and the traditional decontextualized way of sentence-level teaching may not be effective (Levis, 1999).

Shadowing for suprasegmentals

Shadowing is often suggested as one effective teaching and learning method to promote the learning of suprasegmental features within the discourse context. To follow the definition by Luo et al. (2010), the basic skill of shadowing is to immediately follow the utterance produced by the native speaker "as closely as possible." Since learners should remember the speech input and reproduce the original intonation, speed, and stress pattern exactly the same way as the original speaker in real time, it is counted as a highly cognitive action rather than a mere automatic parroting (Hamazah & Miko, 2010). Although this method was originally used for training beginner simultaneous interpreters (Lambert, 1992), it currently has become a widely employed pedagogical technique for improving English pronunciation (Hamazah & Miko, 2010). Furthermore, with regard to shadowing strategy and the teaching of suprasegmentals, Celce-Murcia et al. (2010) hold that shadowing can be considered an effective teaching method for imitating native speech intonation patterns at the discourse level. In this respect, *VoiceTube* can function as one tool for shadowing and ultimately for improving English suprasegmentals in discourse context.

SOFTWARE DESCRIPTION

VoiceTube is a free web-based software that allows practice learning English pronunciation via video resources. As the largest English learning community in Taiwan, it provides over 40,000 videos with 17 different channels of video sources including TED talks, BBC Learning English, CNN Student News, movie clips, music videos, and others with 18 broad categories of topics. One convenient feature of this software is that it has iOS and Android mobile applications. Thus, it is a highly compatible software with current practice of learning English where accessibility and convenience is valued. One helpful feature that distinguishes *VoiceTube* from other seemingly similar resources for authentic English speech such as *TedTalk* or *Podcast* is that this software provides the shadowing tool in its speaking section along with its video section. Thus, it can be counted as a comprehensive system for practicing shadowing technique through listening, speaking and vocabulary.

HOW IT WORKS

Learners can start by choosing their favorite videos from the list. As seen in Figure 1, they can choose the video by channels of interest (17 channels), their levels (basic/intermediate/advanced), and topics (18 categories). Also, they can choose videos by accent (US/UK/AU/Other) and duration.
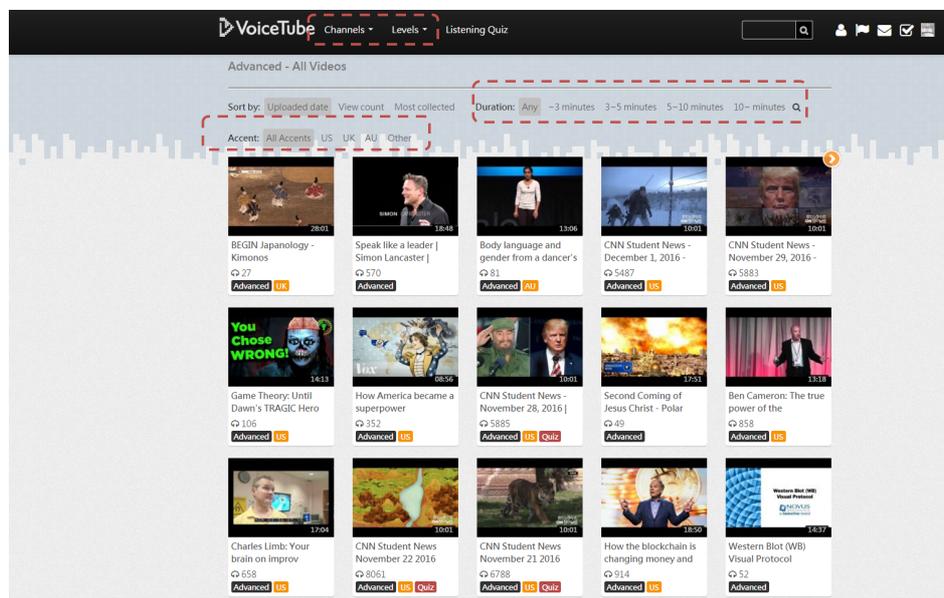

Figure 1. VoiceTube Webpage

The aim of this step is to find the video with the specific accent that each learner prefers to practice. *VoiceTube* provides "other" and "AU" (Australian English) categories in addition to General American English and Received Pronunciation (Levis, 2005). The "Other" category includes the videos of various English varieties such as Philippine English and even includes nonnative speakers of English such as Mexican, Chinese, and Swedish. It should also be noted that learners should select the video which is not too difficult for their levels. Since the ultimate goal of shadowing is to imitate the speech and its suprasegmental features, lack of comprehension should not interrupt the speaking process (Hamada, 2014; Hamazah & Miko, 2010). If learners are in lower levels, it is recommended that they choose videos with shorter durations.

The next step is the listening stage in which learners watch the video in the form of connected speech. As mentioned above, in order to prevent the situation where lack of comprehension skills impede speaking, lower level learners can choose to watch with the help of subtitles provided in English and Chinese and adjust the speed of the speech. As can be seen from Figure 2, they can also refer to the sentence by sentence caption provided right next to the video player. Learners can watch the full video several times or choose to repeat each sentence by clicking the green colored captions.
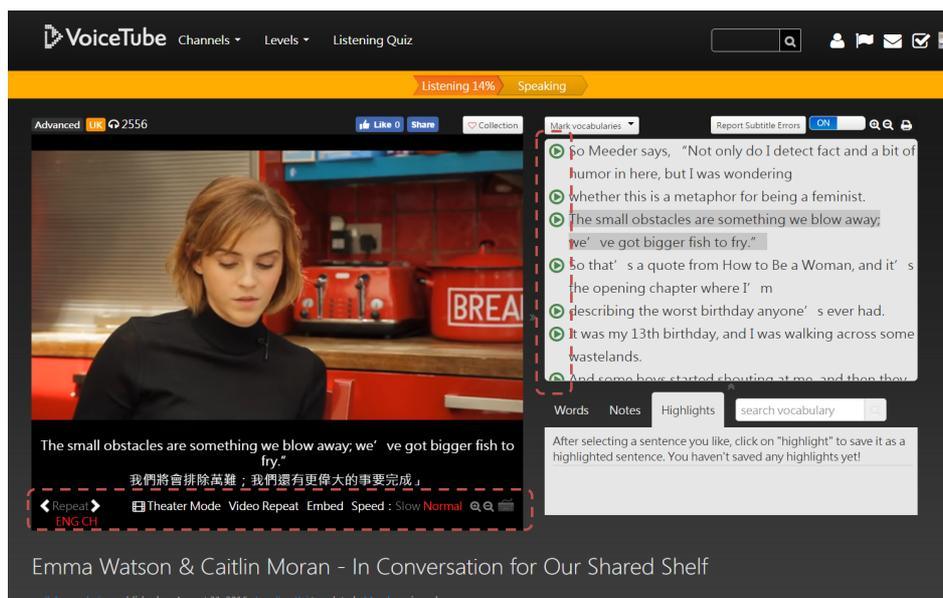
Figure 2. The Listening Stage

The goal of this stage is to listen to the video several times so that learners can become familiar with the speed and the intonation pattern of the speech.

The last step is the speaking stage, or the actual shadowing stage. By clicking on the captions (Figure 3), learners can play the original speech sentence by sentence. Then, they can start imitating the model speech and record it by clicking on the microphone. This process occurs in real time while you are listening to the model speech. If learners stop recording, the player automatically gives them the recording of their speech followed by the original model speech in the video. If further comparison is needed, they can click on the button right to the microphone ("*compare my recording to the video*") and listen to the pair of their own recording and the model speech as much as they want.
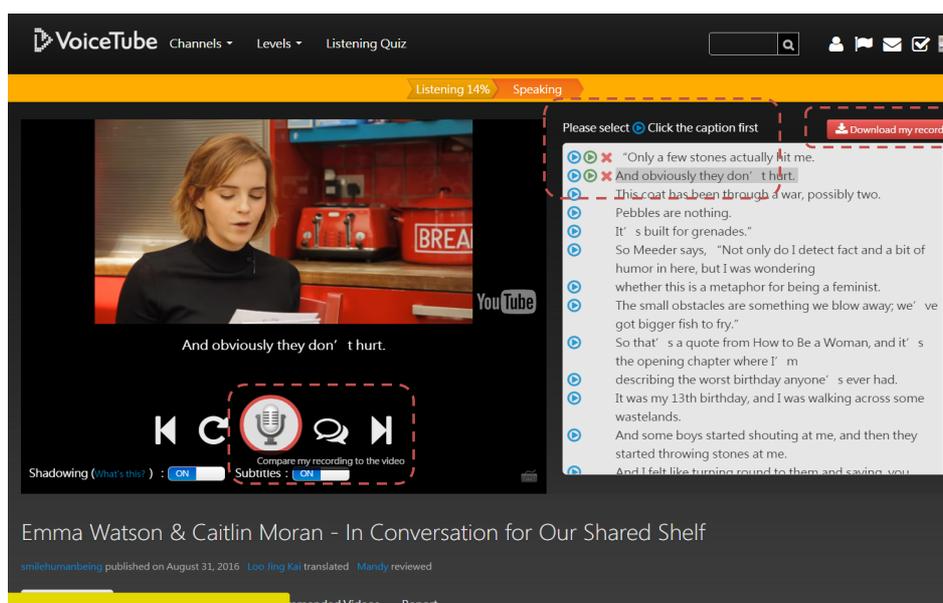

Figure 3. The Shadowing Stage

The aim of this stage is to follow the model speech imitating its intonation, stress patterns, pause, and speed. By comparing their own speech to the model sentence by sentence, learners

can effectively figure out their weaknesses in terms of suprasegmental features and promote learning. They can also download their own speech in mp3 format or keep track of their practice in the account management page.

STRENGTHS

Authentic, discourse-level input

VoiceTube provides a wide array of authentic, communicative, and highly contextualized speech input. The videos can be counted as authentic in that they are the most up-to-date, real life speech about a wide variety of topics. Moreover, the authenticity of this software also comes from the fact that it includes nonnative English speech models along with the native varieties, which is the reality that ESL learners should face in the real life. Considering that current materials for teaching pronunciation are mostly based on *nativeness principle* depending exclusively on American English or Received Pronunciation (RP) (Levis, 2005), those nonnative speech models provided by *VoiceTube* are somewhat inspiring for second language pronunciation teaching. Also, *VoiceTube* is a good source of learning English suprasegmentals since it provides models for discourse-level connected speech. From the previous literature, it is widely accepted that suprasegmental features should be taught at the discourse level rather than in relatively decontextualized and isolated setting such as phrase- or sentence-level teaching (Chun, 2002; Jenkins, 2004; Levis & Pickering, 2004). Thus, all in all, *VoiceTube* is a purposeful tool for learning English pronunciation, especially intonation, because it provides authentic, discourse level input for practice.

Effective functionalities for shadowing

The second advantage of *VoiceTube* comes from its highly effective functionalities as a tool for shadowing. First, it provides a comprehensive system for shadowing composed of listening, speaking, and vocabulary. The specific feature that distinguishes this software from other sources is that it also supports learner production. By way of the recording tool built in the video player, learners can actually record their speech and readily compare it with the model speech. Considering that seemingly similar software such as *TedTalk* only works for listening, *VoiceTube* provides a more effective platform to practice shadowing. Moreover, it provides other convenient and purposeful utilities for promoting shadowing. It is claimed from the previous literature that shadowing is effective after learning target contents, and that learners can use this practice with a lower cognitive burden. (Hamada, 2014; Hamazah & Miko, 2010). Therefore, comprehension of the content should be preceded for successful shadowing of the speech. *VoiceTube* has several functions to enhance learner's comprehension of content. For instance, learners can choose to watch the video with subtitles (either target language or their L1) and captions. They can also adjust the speed of the speech and access the meaning for vocabulary. In this respect, these functionalities can ultimately help prevent the situation where lower level learners' lack of comprehension impedes the shadowing of the speech.

Highly individualized setting

The last strength of this software in terms of the learner fit (Chapelle, 2001) is that it provides a highly individualized setting for learning English pronunciation by giving a wide range of choices for videos by learning purposes, levels, topics, and even target English varieties. It enables the learners to purposefully practice according to their own specific aims. This software is also helpful in that learners can track their learning process, save the videos, save their recordings, and create their own wordbank and notes.

POSSIBILE IMPROVEMENTS

Automatic Speech Recognition (ASR) and feedback

One limitation of *VoiceTube* is that it does not provide acoustic analysis of the learner's recorded speech, or feedback. Current one-way production of recording can be less motivating since it is not interactive. Lower level students, in particular, might need help understanding acoustic features of their own speech compared to the model speech. However, current system does not support speech recognition nor give analysis for the learners. In this respect, ASR can help ESL pronunciation by providing one-to-one feedback (Chiu et al., 2007; Jung, 2011). Therefore, adding speech recognition technology to the current system will help rendering the practice more interactive. If the system offers interaction which resembles the real-life communication more, it will be more motivating for the learners (Alsatuey, 2011).

Expanded levels of shadowing

The second suggestion relates to the way shadowing functions in the software. Currently, it provides a sentence-by-sentence shadowing option only. However, if learners can gradually expand their level of shadowing (e.g. one sentence → multiple sentences → paragraph), it will be more purposeful and meaningful for practicing suprasegmental features in discourse level. In this respect, Levis & Pickering (2004) noted that intonation in discourse is consistently and systematically different from intonation patterns in isolated phrases and sentences. Thus, it would make more sense if the format of shadowing matches the format of the actual speech, which is the connected speech.

Other technical limitations

The last limitation relates to other technical shortcomings of this program. First, the current version only supports English and Chinese subtitles since it was originally developed in Taiwan. Considering that lower level students might need L1 to understand the content of the video, more L1 varieties should be added for wider usability of the software. Also, despite the provision of nonnative model speech in the current system, which is distinct from other softwares, it still has limitations in that they provide relatively restricted varieties of videos for AU and Other category. Thus, I would suggest that they add more sources for nonnative model speech.

REFERENCES

Alsatuey, M. C. (2011). Perceived benefits and drawbacks of synchronous voice-based computer-mediated communication in the foreign language classroom. *Computer-Assisted Language Learning*, *24*(5), 419–432.

Avery, P., Mackin, R., & Ehrlich, S. L. (1992). *Teaching American English pronunciation*. Cary, NC: Oxford University Press.

Chapelle, C. (2001). *Computer applications in second language acquisition: Foundations for teaching, testing, and research*. Cambridge: Cambridge University Press.

Celce-Murcia, M., Brinton, D. M., Goodwin, J. M. & Griner, B. (2010). *Teaching pronunciation: A course book and reference guide* (2nd ed.). New York, NY: Cambridge University Press.

Chiu, T.-L., Liou, H.-C. & Yeh, Y. (2007). A study of web-based oral activities enhanced by automatic speech recognition for EFL college learning. *Computer-Assisted Language Learning*, *20*(3), 209–233.

Chun, D. (2002). *Discourse intonation in L2: From theory and research to practice*. John Benjamins, Philadelphia.

Derwing, T., & Rossiter, M. (2003). The effect of pronunciation instruction on the accuracy, fluency, and complexity of L2 accented speech. *Applied Language Learning, 13*(1), 1-17.

Lambert, S. (1992). Shadowing. *Meta, 37*(2), 263-272.

Levis, J. M. (1999). Intonation in theory and practice, revisited. *TESOL Quarterly*, *33*(1), 37-63.

Levis, J. M. (2005). Changing contexts and shifting paradigms in pronunciation teaching. *TESOL Quarterly, 39*, 369-377.

Levis, J. M., & Pickering, L. (2004). Teaching intonation in discourse using speech Visualization technology. *System*, *32*(4), 505-524.

Lightbown, P. M., & Spada, N. (2006). *How languages are learned*. Oxford: Oxford University Press.

Luo, D., Qiao, Y., Minematsu, N., Yamauchi, Y., & Hirose, K. (2009). Analysis and utilization of MLLR speaker adaptation technique for learners' pronunciation evaluation. In *INTERSPEECH* (pp. 608-611). Brighton, UK.

Hamada, Y. (2014). The effectiveness of pre- and post-shadowing in improving listening comprehension skills. *The Language Teacher, 38*(1), 3-10.

Hamzah Md Omar, M., & Miko, U. (2010). Using 'A shadowing' technique to improve English pronunciation deficient adult Japanese learners: An action research on expatriate Japanese adult learners. *Journal of Asia TEFL*, *7*(2), 199-230

Hsieh, K. T., Dong, D. H., & Wang, L. Y. (2013). A preliminary study of applying shadowing technique to English intonation instruction. *Taiwan Journal of Linguistics*, *11*(2), 43-65.

Jenkins, J. (2004). Research in teaching pronunciation and intonation. *Annual Review of Applied Linguistics*, *24*, 109-125.

Jung, H.-K. (2011). The correction of learner's English pronunciation errors through speech recognition reading program. *Multimedia-Assisted Language Learning*, *14*(3), 291–314.

Pennington, M. C., & Ellis, N. C. (2000). Cantonese speakers' memory for English sentences with prosodic cues. *The Modern Language Journal*, *84*(3), 372-389.

Neri, A., Mich, O., Gerosa, M. & Giuliani, D. (2008). The effectiveness of computer assisted pronunciation training for foreign language learning by children. *Computer-Assisted Language Learning*, *21*(5), 393–408.