Abstract

One trait that sets humans apart from other species is that our communication between species are much more advanced than any other on Earth. However, there are still no agreed upon theory as to how our language and communication evolved due to the lack of evidence. Regardless, humans begin to develop their verbal communication skills very early in life. The development of speech production throughout an individual's life starts from an infant's first babble and is transformed into fully developed speech by the age of five. It is a type of cognitive skill, and thus, we cannot teach it the same way we would teach sciences or history, as cognitive skill learning is the learning of a skill or knowledge that is hard to symbolize. From the moment we are born, we begin to cry as a means of communication. In a few more months we begin to babble, and soon after we form our first words. In just a short few words we begin to produce longer sentence as we attempt to express ourselves deeper. We can observe that a human's first method of communication prior to obtaining the knowledge of language is the verbal method. We can see the importance of verbal communication in a language just from that observation. Despite of that, the way we are taught the verbal component of a language varies wildly. The way the language is taught differs from class to class, and even bigger difference in instructional methods can be seen on the larger scale. Various environmental and cultural factors also affect the way the language is taught and imparted. While the grammatical rules governing the structure of a sentence are usually strictly adhered to by most teachers and students, the same does not apply to the oral component. Even amongst countries that speak the English language natively, there exist a great variation in the accent. When speakers with a vastly different language pick up the English language, their English speech may sound very different from the original. For example, the Japanese language is largely monotonous, but English speeches can be greatly affected by the intonation. This makes the English spoken by most Japanese very hard to be comprehended by other English speakers as the Japanese English lacks the intonation context. These differences in accents results in miscommunication even when communicating using the same language. In addition, there exist many consonants and vowels that are mutually exclusive in both language, and thus, a student who learns to speak English via furigana often end up having a hard time to be understood by non-Japanese English. For example, a Japanese would often pronounce "eight" as "ei-to (エイト), fight as "figh-to (ファイト)", or "the" as "za (ザ). Verbal communication is a major part of a language, but there are not many systems/solutions in the market that caters to self-learning of spoken language. A simple survey on the market place would show that the most popular language learning software focuses on the vocabulary and grammatical aspect. A few of these software contains modules where individual word pronunciations are evaluated as well. However, as mentioned earlier, the intonation of words in an English sentence can affect the meaning, and thus, improving single word pronunciation is inadequate for fluency in language. Speech recognition software can help with pronunciation of longer sentences, but it still does not take into account the intonation. One of the teaching methods that can resolve this problem is Speech Shadowing. It is an experimental technique where a subject repeats speech immediately after hearing it. The process is guided by an instructor who will evaluate the shadowed speech and provide feedback on how to make improvements. However, it is a time consuming method as it requires 1-on-1 tutoring and thus it is not suitable for a large class.

One way we can solve this problem is by applying technology. If we can replace the instructor in terms of evaluating the shadowed speech and provide feedback for improvement, we can greatly increase the adoption rate of this learning method. Depending on the exact technology and technique applied, the system might even reduce training time required. In this paper, we present our approach to utilizing this method for a self-supported learning system and how to utilize technology to improve its efficiency over traditional speech shadowing methods. Using the Cognitive Apprenticeship Model, we describe how the system supports the user in learning via Speech Shadowing. We also explain how the system provide contents and how the system sorts the contents according their difficulty levels. The system would run on a mobile platform to ensure maximum flexibly in self-learning as the user would be able to learn using the system wherever their smartphone/tablet goes. The user management system would also ensure that users are given an approximation of their current progress so that they can self-motivate and also not attempt speeches that are too far above their proficiency level and in turn get discouraged. The system also provides several forms of scaffolding (support) to help weaker users improve. The system uses self-evaluation and it provides adequate support for users to accurately self-evaluate. This is in the form of audio waveforms. By comparing certain traits in 2 audio waveform, users can abstract valuable information from the raw data. An algorithm was defined to evaluate the user's proficiency level based on the data gathered by the system during the self-evaluation process. This data is also used by the system to determine the types of support needed by the user should he/she attempt speeches at levels that are different that his current level. The paper also describes the system development process on the Android platform and examples of a standard use-case flow with the system interface. At the end, a case study was conducted to validate the effectiveness of the system described. We find that the implemented system can provide coaching similar to that of a human instructor. In the final section of the paper, future works and improvements on the system is describe.