Online Speech-to-Text for Speech Training: The Perspective of Learners of English

Joao Carlos Koch, Jr.
Sapporo Gakuin University

This project analysed the effectiveness of a speech-to-text tool for language learning. Speech-to-text has existed for a while, and its main purpose is to allow users to dictate text in their native language with accuracy. With most tools, accuracy is enhanced through trainability; that is, the tool identifies the user's pronunciation patterns and adapts itself to them. However, this is not desirable for language-learning situations because it would produce unrealistic, flawed feedback with decreased usefulness for the language learner. This research was conducted by choosing one tool suited to learners' needs, which participants then used in an e-learning situation, with focus on speech training and phonemic and phonetic awareness-raising processes, guided by the tutor. Participants reflected on and evaluated the experience and the chosen tool. Results indicate that the experience is valuable and promising, although the tutor role is paramount in both providing feedback as well as promoting learner autonomy.

　この研究では、音声・テキスト変換ツールを用いた語学学習法について述べる。音声・テキスト変換は以前より存在しており、これらのツールの主な目的は利用者の母語での口述を文章化するためのものである。多くのツールは訓練可能であり正確性を強化できる、つまり、ツールはユーザーの発音パターンを識別し、そのパターンに適応する。しかし、この機能は語学学習においては望ましくない。なぜなら事実とは違うフィードバックを与えてしまうからであり、これは語学学習者には有用性が低い。この研究では学習者に適したツールを選び、参加者はこれをeラーニングのコースで使用した。参加者は講師の指導によりスピーチ訓練と英語の発音を重点的に取り組み、その後この学習体験とツールを評価した。この研究で、この学習法は効果があり将来性があるものだということが明らかになった。しかしフィードバックを与えたり、学習者が自主的に学習できるようにする役割として、講師の存在は重要である。

THE AIM of this project was to investigate the usefulness and relevance of speech-to-text in general, (and one tool in particular) for learners of English, from the learners' points-of-view. In general, learners of English who live in their home country have limited opportunities to practise their language skills, particularly oral communication, and only in a fraction of these instances are they able to receive feedback on their speech. In a classroom setting, for instance, opportunities for individual feedback are limited, and often teachers adopt a one-size-fits-all approach when it comes to language feedback. Opportunities for teachers to address specific problems in a learner's pronunciation are often difficult to come by. On the other hand, learners often attain high levels of proficiency and fluency in a language without necessarily having good pronunciation. Assessment of speaking, when present, often places more emphasis on grammatical accuracy and fluency because an incorrect choice of words is more likely to cause confusion than a misplaced vowel or consonant.

The process of training learners to interpret the results from speech-to-text tools involves raising their awareness of areas of phonology, which in turn can potentially equip learners with the knowledge needed to monitor their own speech and improve their overall phonemic and phonetic accuracy. The simple fact that learners may have to repeat sentences several times when practicing with a speech-to-text tool means they will have more opportunities to practise and perfect their own speech in terms of not only accuracy but also fluency.

In spite of past research and investigations, as well as recent enhancements in both the technology and its availability, there have been limited research investigations into the applicability or feasibility of speech-to-text tools in second language learning. Moreover, past studies focused on results rather than learner perception, and I believe the latter is a valid area to explore. In this study I sought to establish to what extent learners become more aware of their own pronunciation as well as the features of English speech articulation. This paper reports on a case study conducted to find out how learners perceived the helpfulness of a selected speech-to-text tool in second language speech training.

Literature Review

Speech-to-text, also sometimes referred to as *speech recognition* or *voice recognition* (these also include voice instruction, without necessarily producing text)*,* or simply *dictation*, is a type of technology that has been available for some time to most computer users. The description may vary slightly depending on the source, but speech recognition software is generally defined as technology that recognises speech and uses it to connect humans and computers (The National Center for Technology Innovation [NCTI], 2010), or more generally, as programs that translate spoken words into text (Wikipedia, n.d.). In this paper, I will refer to this as *speech-to-text*.

Speech-to-text tools were originally created with increased productivity in mind. For instance, they can be very helpful in a setting where the user is unable to physically type or to eliminate the need for a second person to write down what one is dictating. Speech-to-text was not originally intended to be used by users speaking in a language they are not proficient in, mainly due to likely differences in their pronunciation and speech patterns compared to native or near-native speakers.

Although in this study I aimed to explore the feasibility of speech-to-text tools for language learners' speech training, such practice should be approached with caution. There are mixed results, opinions, and conclusions regarding speech-to-text tools in general and their use by nonnative speakers. Some authors have recognised the potential of this technology for use in second language learning (Coniam, 1998; eflclassroom.com, 2012; Gorban, 2012; Hancock, 2013; Myers, 1999; Neri, Chucchiarini, & Strik, 2003; Price, 1998). Others believe the technology still needs improvements before it can be used in this setting (Derwing, Munro, & Carbonaro, 2000; Schneiderman, 2000). Furthermore, even authors who were generally positive about the use of the technology by nonnative speakers tended to be cautious about it. For example, Price (1998) stated that speech recognition technologies were generally not intended for classroom use, and educators often lacked the background necessary to use the technology efficiently and lacked awareness of its limitations. Myers (1999) warned that the tool, although useful for speech training, does not address other, more specific features of human interaction—metalanguage, such as gestures or vocal cues.

One should bear in mind the possible implications (in terms of potentially reduced accuracy) of the use of speech-to-text by nonnative speakers. According to Coniam (1998), makers of speech recognition technologies "make great claims for the accuracy of the software" (p. 8). Based on the results from his experiments with native and nonnative speakers, Coniam explained that although one maker's promotional literature claims "an accuracy rate of 95% or better," these claims are "far from being achieved" (p. 8). Native speaker scores were substantially below the claimed 95% and

nonnative speakers scored even lower (pp. 10-11). This difference between the recognition of native and nonnative speakers' speech is consistent with results found in at least one other experiment. Derwing et al. (2000) claimed that the software used in their study showed poorer performance at recognizing nonnative speakers' speech, which, according to the authors, can have "important implications for pedagogical uses of the software" (p. 593). In spite of this, Myers (1999) defended the speech-to-text tool as a useful approach to the development of skills and strategies related to English speaking, as well as for autonomous learning, even when the tool is unable to recognize an utterance.

Speech-to-Text Tools

Although research on speech-to-text has made a great deal of progress in recent years, there are few studies that investigate the use of this technology for the purpose of language learning from the viewpoint of language learners. Nevertheless, as more studies on speech-to-text are being conducted, the actual capabilities of the technology have increased, and many tools now offer a range of features that go beyond the basic speech-to-text capability. Features that are relevant for this project are *trainability* (a narrow term, referring to a specific feature present in some tools) and *accessibility* (a broader term, encompassing characteristics related to the access and use of the tool).

Trainability

With this feature available, users can train the software to recognise their speech, including their accent, therefore increasing the tool's accuracy. The aim of this feature is to enable the tool to increase users' productivity in the long term, making the tool more accurate and responsive to the user's speech and less prone to misrecognition. The fundamental purpose for training (to enable the tool to transcribe the user's speech more accurately by overriding poten-

tial inaccuracies and building a database that includes the user's mispronounced words) would undermine somewhat the purpose of this study, which was to provide realistic feedback about the user's speech without such facilitators.

Accessibility

This refers to online or desktop access to speech-to-text tools. The advantage of thee easily accessible tools, compared to other tools that require installation, is that users can access online tools from any machine that supports them. Therefore, they are not bound to a limited number of machines and do not need to go through any installation or set-up process. However, the accuracy and lexical range of these online tools are arguably inferior to those of tools that require installation (mainly due to the fact that the latter can often be trained). Table 1 summarises the main features of online and offline tools.

Table 1. Categorization and Features of Speech-to-Text Tools

Factors	Online speech-to-text tools	Offline speech-to-text tools
Charac-teristics	No installation required Multiple-machine use Free	Trainable Superior accuracy Expandable database Additional features beyond basic speech-to-text functionality
Limita-tions	Database not expandable Limited accuracy Few features beyond speech-to-text functionality Nontrainable	Installation required Paid licence (if third party) Single machine use (need to be installed on every machine before use)

For the purposes of this study I chose to use Talktyper (talktyper.com), for the following reasons:

- It is ready for use on any computer, by any user, and computers can be shared.
- It is free.
- It is not trainable and its database is nonexpandable.

The only drawbacks are that an Internet connection must be available and that the accuracy of online tools is inferior to that of their desktop counterparts.

Procedure

The participants were six learners of English who volunteered to take part in the study and who came from a variety of backgrounds. All spoke Japanese as their first language and all were female. Full, individual profiles can be found in Appendix B. They were given instructions through a webpage designed for the course and they posted their findings in a discussion forum. To avoid any feelings of embarrassment, each participant was isolated from others in the online environment, so that their information and contributions to forums were hidden from all except the tutor.

The project was divided into five steps, with each step having a higher degree of complexity than the previous step (see Table 2).

The first step was to orientate participants on the course, and to list the environmental and technological requirements in order to progress to the following steps. The second step was to offer an introduction to Talktyper, and to enable participants to become familiar with its basic functionality. The third step was similar to the second, but this time participants were required to dictate a full text. They were given suggestions as to what strategy to choose for this undertaking, but how they performed the task was left up to them. At the end of this step, they were required to submit their dictated version and feedback was given.

This feedback was an analysis of their dictated text in comparison with the original text (which was the same for everyone). The analysis compared mismatches between the two and provided an explanation for possible sources of interference, (e.g., unstressed sounds not heard by the machine, a weak microphone, or inaccurate pronunciation by the participant). The rationale for this approach was to equip participants with a set of skills that would enable them to take an active role in performing a similar analysis for the next step in the project. Even partial success in this analysis would mean the participant had raised her awareness about features of English speech and pronunciation.

The fourth step required participants to use the knowledge gained thus far to dictate a short text of their own creation (language feedback was given on their writing prior to the dictation by the teacher–researcher). The resulting text from this dictation was then analysed by the participant together with the teacher. Suggested reasons for mismatches between their original text and the dictated text were elicited from participants, with the teacher assuming the role of facilitator. The last step required learners to give a short, oral presentation of their written text in a one-to-one setting with the teacher.

Table 2. Speech-to-Text Project Staging

Step	Activities
1	Orientation about the course
	Check of environmental and technological requirements to perform the tasks
2	Introduction to Talktyper (to enable participants to become familiar with the basic functionality of the tool)
	Dictation of three selected statements
	Short analysis of the dictation, checking for accuracy

Step	Activities
3	Participants given a few different suggestions as to what strategy to choose for dictation
	Dictation of a preselected short text (~100 words)
	Feedback on the dictated text
	Selection of a short text or video to summarise
4	Dictation of summary
	Analysis of dictated text (with and without teacher assistance)
5	Short oral presentation to teacher of the content of selected text or video
	Discussion

Reflections on the Project

Some of the primary factors that likely caused problems in this research are not related to the tool's performance but rather to the limited experience of learners in the use of speech-to-text tools. Such limitations were not anticipated when the project was designed. Neri et al. (2003) noted that a number of issues identified in the publications they analysed were due to factors not directly related to speech recognition technology, but that these issues were attributed to the tools because of little familiarity by users with this kind of technology. Due to the limited duration of the current project, there was insufficient exposure to eliminate or even minimize these problems.

Another issue was related to the relatively low level of commitment by participants. There seemed to be a tendency to want to finish the steps in order to reach the end of the project, instead of concentrating on understanding the process. This led participants to place less emphasis on the course content. I believe this happened (at least partly) because participants were invited, rather than required, to take part in the project. Their low motivation was not unexpected because there were no major benefits to participating, nor was there any perceived necessity to use the tool prior to, or even during, the course by many participants.

In terms of recognizing the participants' pronunciation, I believe the tool's performance was satisfactory and did not impede participants from obtaining useful feedback. In a few instances, learners turned in dictated text that indicated Talktyper had performed poorly. However, this was remedied by a second attempt, after changes were made to the tool's settings (e.g., changing the language variety from American English to British English), or the physical setting (e.g., moving to a quieter area or using an external microphone). The same was found by Neri et al. (2003), who stated that the recognition performance of nonnative speech by the speech recognition tool seemed to satisfy most teachers.

I adopted an inductive approach in the project. That is, participants were required from the early stages to identify and analyse potential areas for improvement in their own dictated version of a text. However, because participants placed emphasis on finishing the steps rather than understanding the project, the method I had chosen to deliver feedback (i.e., individualised responses) was not efficient. Although this can be beneficial for learners who are more motivated and willing to spend time to understand the subject and decipher the text from Talktyper, the approach did not have this effect on the participants in this project.

Account also needs to be taken of the differences between speaking to write and speaking in conversation. Speaking to write requires different skills than speaking in conversation, and the NCTI (2010) suggested that learners should be made aware of these differences. This difference was not stressed in this project, and participants worked on speaking to text only. Another limitation was that the project was not designed to elicit spontaneous speech. Participants worked entirely with speech produced by reading a text, so the use of Talktyper for spontaneous speech cannot be inferred from the results. Schneiderman (2000) wrote that free

speech interferes with other cognitive functions. Perhaps tools like Talktyper are still not adequate to work with this speech and generate useful feedback for learners who are nonnative speakers, due to their possibly irregular speech patterns. More research is needed into this problem.

Analysis and Evaluation of Participants' Feedback

This section describes the participants' evaluation and feedback of their experience using Talktyper. Two instruments were used to evaluate participants' perceptions about Talktyper and to gather data for this evaluation: a questionnaire and an interview.

The questionnaire (see Appendix A) was divided into three parts:

- Part A was to gauge participants' feelings about the tool and their experience using it. Each question was answered on a Likert scale (from 1 to 7) and a space was provided for participants to justify their choices.

- Part B contained a list of positive and negative descriptors; participants selected those that most accurately described the tool in their view.

- Part C was to uncover specific pieces of information and evidence of raised awareness and learning that took place during the course of the programme, through open-ended questions.

The interviews were conducted with each participant individually, in the form of short conversations of between 10 and 15 minutes in length, and were aimed at extending and expanding on the questionnaire responses. Each interview consisted of the participant explaining her questionnaire answers further, elaborating on any answer for which the meaning was not entirely clear to the researcher, justifying her answers, and double-checking her answers as well as her understanding of the wording in the questions if necessary.

The combined data collected from both instruments is shown and described below. Points 1 to 8 analyse results from part A; point 9, from part B; and points 10 to 12, from part C. For a collection of contributions from participants that address these points, see Appendix C.

Summary of Results

Overall, participants agreed that Talktyper was either useful or very useful. There were mixed results regarding participants' perceived ability to use the software on a regular basis. Most participants felt that the tool was pleasant to use, which could have positive impact on their motivation. However, this joy factor was closely linked to the tool's accuracy at a given time, and low accuracy or performance could significantly impact users' enjoyment and therefore, motivation. All participants demonstrated at least some raised awareness of English pronunciation, and a few managed to give more detailed explanations about their learning. Most participants acknowledged that Talktyper is potentially very useful for language learning and speech training, in spite of the reservations some of them had regarding the tool's performance. A number of participants were less optimistic about using the tool without teacher assistance. A few participants, however, demonstrated higher confidence in their ability to use the tool autonomously, but they too stated that their productivity would be limited. The likelihood of participants recommending Talktyper to friends or colleagues ranged from unlikely to very likely. In general, participants who encountered more problems were less likely to recommend this tool to their peers.

Although all participants agreed that Talktyper would benefit from at least a small number of improvements, they felt that the tool was efficient enough. Participants generally praised Talktyper, its performance, and their overall experience. Positive descriptions that were most frequently used were *useful, enjoyable* (4 occurrences each), *helpful, simple, user-friendly*, and *easy to operate* (3 occurrences each). The most commonly used negative descriptions were *frustrat-*

ing (3 occurrences), *challenging*, *stressful*, and *confusing* (2 occurrences each). Participants' understanding of the project's aims varied, but all remarks mentioned an aspect of learning in some way. Most participants were able to visualize situations that are likely to benefit from Talktyper. Although the majority of participants claimed that they would need a teacher to help them with interpreting data obtained from Talktyper's dictated texts, all of them were able to give examples of one or more features of English pronunciation and how this knowledge, together with Talktyper's feedback, helped them to notice areas that needed improvement.

Discussion

From participants' written justifications as well as their interview responses, we can see that there were a number of technological problems.

It is evident that the participants' perceptions of the tool were heavily influenced by their emotions and feelings. For instance, when choosing words to describe the tool and their experience, descriptors like *fast, slow, inaccurate,* and *complex* were not chosen by any participants, and other words (that also describe features other than feelings or emotions) such as *quick, inadequate, glitchy, unreliable,* or *inaccurate* were chosen only once each. Participant motivation (before, during, and after the process of dictation) is an important aspect to factor in when planning an activity involving Talktyper. That the participants chose words and phrases that describe feelings and emotions offers an important insight into the learners' attitude towards an unknown task. It is natural that learners, without the expertise of a teacher, are likely to make conclusions that are based on what they are experiencing and how they feel about it.

Participants' limited exposure to the tool perhaps reduced their ability to analyse the tool or comment on its characteristics in more detail. This may be a factor that influenced the higher occurrence of descriptions of feelings or moods in comparison with comments about the tool.

Conclusion

This study examined the use of speech-to-text tools for language learning, but the information gathered was limited due to the short duration of the programme. Although my primary aim was to focus on participants' impressions of this technology for the purpose of speech training, a longer programme (possibly using a variety of different tools, including mobile devices) could potentially result in a deeper understanding of the effects of this tool on learning and pronunciation awareness, and the value of the tool for the language learner.

Considering the overall positive feedback from participants and the fact that all participants were able to display new understanding of pronunciation awareness as a result of the project, I believe the use of speech-to-text tools would be a welcome addition to the repertoire of applications used in language learning. Another desirable learner behaviour that speech-to-text tools allow for, due to their wide availability, is autonomous learning. Technology can be used to provide learners with more opportunities for speech training, but without guidance the benefits of this practice are likely to be limited. Currently, the only nonhuman source of speech feedback on the Internet is English Central [englishcentral.com], and even this website is limited in terms of content compared with the potentially unlimited scope of combining a tool like Talktyper with teacher guidance and feedback.

Therefore, although autonomous, self-directed learning is often a goal, it is important that the use of speech-to-text tools should happen in a more controlled setting, at least during the initial phases of implementation, with (a) active tutor guidance (to help learners use the tool), (b) monitoring (to ensure learners are on task and to troubleshoot problems), and (c) feedback (to negotiate the meaning

of the chosen tool's feedback and its value to learners' pronunciation). It is equally important that tutors possess a solid understanding of the phonetics and phonemics of both the target language and learners' L1, so as to be able to provide learners with informed and precise feedback. It is hoped that a long-term result of this monitoring and awareness-raising process might be that learners can be weaned off the support of a tutor and trained to autonomously interpret the feedback from their dictated texts.

Bio Data

Joao Carlos Koch, Jr. trained and taught in New Zealand, before marrying and settling down in Hokkaido, where he teaches and does research. He holds a Trinity DipTESOL and is currently completing an MA in TESOL. His special interests are learner motivation and autonomy, teacher development, educational technology, and in-depth interviewing.

References

Coniam, D. (1998). The use of speech recognition software as an English language oral assessment study: An exploratory study. *CALICO, 15*(4), 7-23.

Derwing, T. M., Munro, M. J., & Carbonaro, M. (2000). Does popular speech recognition software work with ESL speech? *TESOL Quarterly, 34*, 592-603.

Eflclassroom.com. (2012). *Alan Schwartz: English Central* [Interview]. Retrieved from http://www.eflclassroom.com/eltinterviews/?tag=speech-recognition

Gorban, J. (2012). *Using dictation software to study foreign languages*. Retrieved from http://gorban.org/post/26549258297/using-dictation-software-to-study-foreign-languages

Hancock, M. (2013). *Robin Walker on technology in pronunciation teaching*. Retrieved from http://hancockmcdonald.com/blog/robin-walker-technology-pronunciation-teaching

Myers, M. J. (1999). Voice recognition software and a hand-held translation machine for second language learning. *CALL, 13*(1), 29-41.

The National Center for Technology Innovation (NCTI). (2010). *Speech recognition for learning*. Retrieved from http://www.brainline.org/content/2010/12/speech-recognition-for-learning_pageall.html

Neri, A., Chucchiarini, C., & Strik, W. (2003). Automatic speech recognition for second language learning: How and why it actually works. *Proceedings of International Congresses of Phonetic Sciences.* Universitat Autònoma de Barcelona, 1157-1160. Retrieved from http://hstrik.ruhosting.nl/wordpress/wp-content/uploads/2013/02/a102.pdf

Price, P. (1998). How can speech technology replicate and complement good language teachers to help people learn language? In *Proceedings of Speech Technology in Language Learning* (pp. 81-86). Marholmen, Sweden: STiLL.

Schneiderman, B. (2000). The limits of speech recognition. *Communications of the ACM, 43*(9), 63-65. Retrieved from http://www.cs.umd.edu/~ben/p63-shneidermanSept2000CACMf.pdf

Speech recognition. (n.d.). In *Wikipedia*. Retrieved from http://en.wikipedia.org/wiki/Speech_recognition

Appendix A
Questionnaire
Part A

For questions 1 to 8, please mark the space on the scale that best represents your answer in relation to the left and right ends. When possible, give reasons for your choices. (Likert scale of 1-7 included for each question.)

1. After this experience, how useful do you believe speech-to-text software has become for you?

2. How easy it is for you to use speech-to-text software regularly?

3. How much do you enjoy using speech-to-text software?

4. How much were you able to learn about English pronunciation through the software's feedback (text) on your speech?

5. How beneficial are speech-to-text tools for language learning, particularly for speech training?

6. How competent do you feel to interpret speech-to-text software feedback without the help of a teacher?

7. How likely are you to recommend this tool to your friends or colleagues?

8. How much improvement does this tool need?

Part B

Based on your experience and future expectations, choose the descriptions that best describe speech-to-text tools for speech training, according to your perception:

- Valuable
- Enjoyable
- Rewarding
- Fast
- Useful
- Helpful
- Pleasurable
- Easy to operate
- User-friendly
- Quick
- Fun
- Accurate
- Simple
- Valuable
- Enjoyable
- Rewarding
- Fast

- Useful
- Helpful

Part C

Please answer the following questions:

1. Explain in your own words what you believe this project aimed to find out.

2. Are there situations where you would use English and speech-to-text would be useful? If yes, please name a few. If no, please state so.

Think about the things you've learnt about using speech-to-text software. Describe what you've learnt about:

3. The technology itself (including its strengths and weaknesses)

4. The tool's feedback

5. Your own mistakes

Consider your own speech and pronunciation. Explain briefly what you've learnt in terms of (if you're not sure, or don't have to add anything to a particular question, please write so):

6. Accuracy in pronouncing the consonants of English

7. Accuracy in pronouncing the vowels of English

8. Running words together (i.e., no pause between two or more words)

9. Missed out sounds (i.e., vowels or consonants that are written, but not pronounced)

10. Strong and weak vowels

Appendix B
Participant Profiles

Participant A
- Gender: female
- Age: mid 30s
- Occupation: university teacher/researcher
- Purpose for learning English: for research, international conferences, conversation with foreign colleagues
- Nationality/L1: Japanese
- Estimated English level (CEFR): B1

Participant B
- Gender: female
- Age: early 20s
- Occupation: graduate student
- Purpose for learning English: for future job, research and study opportunities
- Nationality/L1: Japanese
- Estimated English level (CEFR): B2/A1

Participant C
- Gender: female
- Age: late 20s
- Occupation: graduate student
- Purpose for learning English: for future job, research and study opportunities
- Nationality/L1: Japanese
- Estimated English level (CEFR): B2

Participant D
- Gender: female
- Age: early 30s
- Occupation: office worker
- Purpose for learning English: for leisure, travelling and conversation
- Nationality/L1: Japanese
- Estimated English level (CEFR): B1

Participant E
- Gender: female
- Age: mid 20s
- Occupation: pharmacist
- Purpose for learning English: for leisure and future job opportunities
- Nationality/L1: Japanese
- Estimated English level (CEFR): B1

Participant F
- Gender: female
- Age: late 20s
- Occupation: therapist/researcher
- Purpose for learning English: for leisure, research, international conferences
- Nationality/L1: Japanese
- Estimated English level (CEFR): B1/B2

Appendix C
Categorised Summary of Participants' Contributions to the Questionnaire and Oral Interview

1. On the tool's usefulness

PA: When talk typer represented my correct answers, I was happy and motivated by the system, but when it showed incorrect answers, I didn't get enough information for improving (and I felt a little sad and irritated).

PB: The software made me careful about my pronunciation and to read texts more appropriately because it has great recognition ability.

PC: This tool is useful to practice speak English. Especially it is good that computer pronounces the words in correct pronunciation (accent). But this tool has microphone problem. This problem sometimes discourage me.

PE: One good point is that I can assure my pronunciation is almost correct or not most of the time.

2. Ease of use

PB: It's really interesting because before I used this software I imagined dictation software didn't work so well. However, since I used this tool, I realised it worked almost perfectly.

I used this almost without any problems. However, the software would be nicer if it had a better text editing tool. I sometimes felt difficulties to rewrite the wrong recognitions.

PA: It has a friendly interface, so it was easy to use it for me.

PD: I could use this whenever I wanted.

PC: I have a laptop at home, so it is easy to use this tool.

3. Enjoyment factor

PE: It is difficult use completely manage this software. But when I can understand why it recognised something in a particular way, it's a good tool to progress my pronunciation.

PC: If the sentences I spoke were recognised by talk typer, I could enjoy it. But sometimes I had many mistakes and many difficulties to input the correct words.

I study really hard to speak English accurately, so it is stressful when I find too many mismatches between Talktyper's feedback and what I actually said.

PA: When it showed my spoken words correctly, I was happy. I feel good when it reacts to my voice, when results match what I said. But often TalkTyper showed a different sentence so I wasn't in a good mood.

PF: I think first 10 or 20 mins, I was able to enjoy this tool because it's like a game.

To complete this task, I often needed quite a long time. And, I usually had to accept quite a lot of negative feedback to my speech from this tool. That's why it's difficult for me to keep my motivation.

Usually TalkTyper didn't reconise my pronunciation. Usually I had to try several times even very easy words. So 20 minutes later I got tired, both physically and mentally. My mouth was tired and sore.

PD: When Talktyper makes a sentence that is the same one I speak, I feel really happy!

It's fun because I can speak English in any way I want. I'm usually shy to talk to people, but I have no trouble with speaking with a computer. Computers don't respond, but this I'm OK with it.

4. Perceived learning/awareness

PF: Of course when I chat I usually don't care about my mouth, but when using TalkTyper, Id like to improve my pronunciation. So that's why I really carefully moved my mouth and tongue.

If TalkTyper gives me very negative feedback, I can understand why it happened.

PD: I could find out some of my weaknesses, but I think I still need a teacher's help.

PA: I couldn't understand the reason why I got English words with the correct pronunciation or not. I just knew whether it was correct or incorrect from Talktyper, but I didn't know why.

PC: I should use this tool more. I feel it may be difficult to be aware what the wrong points were in my speaking only by myself. But the feedback from the tutor was really beneficial for me.

PB: I was surely careful about my pronunciation, but talking directly with native speakers and copying how they pronounce seem to be better for me to learn foreign languages.

5. Benefits for language learning

PC: Although there seemed to be a problem with my microphone, it is very good chance to speak English very carefully, one to one.

I feel the chances to speak English increased through this project, for me.

PE: It's good as one way to learn English.

PA: When I have to do speeches from a script, the tool is good for practicing.

I cannot learn about vocabulary or grammar because of Talktyper. But for how to use my voice or pronunciation, it is useful, I think.

When we use a dictionary, like a book, old-fashioned, there's only the word or sentences, sometimes the instruction to pronounce it. But we cannot know if our pronunciation is correct. If we use Talktyper, we can check our pronunciation.

PB: The software has great benefits when studying on my own, especially for reviewing.

PD: Speaking is the most difficult and the most important skill, I think. This tool can be used to train it!

I think it's useful because I can speak more. I can say whatever I want and get feedback.

6. Autonomy in using the tool

PC: Maybe now I still can't use it by myself, but when I get more speaking skills and listening skills, I will be able to use it.

It is still difficult to find my trend of mistakes only by myself.

When we look into Talktyper's feedback together, I can understand it. I feel I need to have someone to guide me and help me interpret the tool's feedback.

PB: I have little knowledge about syllables, so all comments from my teacher were totally new to me.

I'd be able to use it both by myself and with a teacher's help. But I strongly feel I need a teacher's help, because there's no advice on how to improve my English after I use this software after I dictate. But with teacher's help, I think I'm able to improve my pronunciation after using the software. It should be combined, self-study but aided by a teacher.

PA: I can't get enough information [from TalkTyper] and use it for my improve-ment.

PE: It's difficult for me to use it by myself.

PD: This tool is very useful and I can practice whenever I want, but I think I'm worried about my English if nobody helps me, so I still need a teacher's help.

I think I have to know how to pronounce words, that's why I think I need a teacher's help. I can use the tool as a supplement, or as a main resource, as long as I get help.

7. Considerations on recommending the tool

PB: Most of my friends have little time to study or practise English. So, I think this kind of practice and feedback would be useful for them. But without a teacher, I think they may have limited benefits because of their limited knowledge about English pronunciation.

PE: This is one of the good ways to practice peaking. But this is not enough to cover a full speaking training. It can't be used alone and I would not recommend it to everyone.

PD: I don't need to be shy for speaking with TalkTyper and I can practice anytime!

This tool is useful to practise general conversation. But if I want to learn English for a more specific purpose, like business English, I think I should go to an English class, it's better to talk to a person.

PC: It is useful to practice English for presentations or speeches.

PF: For example I have friends or colleagues who want to learn English. But usually they study English very passively. I'm not sure they can keep their motivation, I guess its too tough for them. If I know a person with the same ability as me, I will recommend this tool.

PA: I would like to recommend it for persons who want to practice for speeches, but I don't recommend it for person who want to practice for English conversation; whether I recommend or not depends on the aim of users.

8. Impressions on the technology/problems with the tool/suggested improvements

PF: I often couldn't understand why this tool didn't recognise my speech. If I speak in Japanese, I can tell which are the main factors on mistakes, and whether it's because of my pronunciation or the software system. In English, I wasn't sure why till the teacher explained that reason.

It's hard for me not to get depressed if it give me bad feedback, for instance, when it doesn't identify more than 80% of my speech. It was difficult to keep my motivation when this tool didn't identify most (or all) of my sounds.

Everyone can use it without any charge if they have a computer and it can be connected to the internet. It allows users to choose between American English or UK English.

PE: I'm not sure about improvements needed. The tool's performance maybe depends on the person using it.

Typing is easier than speaking.

It's better to talk to people instead of computer. I need someone to help me, and a computer is not enough. I can study grammar and words by myself, but for pronunciation, I need someone to judge if it's correct or not. So this software helps me to learn how to speak and how to pronounce words but finally we need a human tutor.

PA: I didn't know about such software, I was very surprised.

Talktyper is too sensitive. If I change the volume or my voice, the result will be different. I cannot tell the difference between the first or second times, but Talktyper shows different results. If I say a word in two occasions, the result may be different, but I can't tell why.

The tool gave me feedback but it showed different results according to changes in my environment such as volume of voice and breath.

I think people who use this tool regularly may get more information for improvement.

PC: I found difficulties that I think are from a problem on my microphone. Maybe with a better setup I would have been in a better position.

PD: Teachers should find out their students' weakness and give homework to them. I also would like to take tests to know my level sometimes.

When I speak slowly, it starts to show sentences before I can finish speaking.

9. Tool and experience – descriptors (the number following each descriptor represents how many participants chose each descriptor)

Valuable 3	Confusing 2	Rewarding 2	Fast 0
Stressful 2	Useful 4	Unreliable 1	Pleasurable 2
Easy to operate 3	Simple 3	Glitchy 1	Slow 0
Time-consuming 1	Enjoyable 4	Boring 0	Inaccurate 0
Helpful 3	Challenging 2	Intimidating 0	User-friendly 3
Frustrating 3	Inadequate 1	Quick 1	Fun 2
Accurate 1	Complex 0	Tiring 1	

10. The aims of the programme

PA: I think the purpose of this project is to study how English learners keep their studying with a PC.

PF: I think the aim is to find how this tool can help a student when they study English pronunciation and speaking, especially by themselves.

PD: The aim is to improve our speed when speaking English. I think I can speak faster if I speak often, and this tool helps me definitely.

PC: The aim is to find out the correct pronunciation, intonation [and stress] of English (some words should be spoken strongly but other words are weak), and patterns of my mistakes.

11. Situations to use TalkTyper

PF: [I can use this tool] when I study English about pronunciation and speaking.

I think I can use this tool when I study English by myself. Usually I have one lesson per week. So I have six days a week and I can study English on these six days. And TalkTyper is one method to study English.

PC: [I can use this tool] to practice for presentations in English.

PB: It would be useful for self-study. Perhaps before an oral presentation, to rehearse.

PE: I can take notice when I use English words that I studied through this software.

PA: When we do speech from a written script, [TalkTyper] will be good trainer. But it won't be a nice trainer for English conversation, that is, free speaking without a script.

Talktyper is a very cold teacher, I think. This is fine, because the decision level doesn't change. Sometimes human teachers say 'oh, it's OK', even when students say something that is not OK… or maybe it is OK. But Talktyper only game me feedback, answers, without giving information about how to improve my pronunciation.

12. Examples of pronunciation awareness

Consonants:

PC: fat, hat, foot, hoot - ban, van, boy, voice – differences between [f, h] and [v, b]

PF: I couldn't pronounce /l/ and /r/ sounds as different sounds.

I sometimes had to exaggerate my mouth articulation. For example when I say the /f/ sound, I know I have to bite my mouth.

PE: I learnt about the /r/ and /l/ pronunciation. For example, *fruit* and *flute*. These sounds are different from my pronunciation. Sometimes I need an explanation about what's different between the correct pronunciation and my pronunciation.

Vowels:

PF: It is so difficult to pronounce *a*. English has some sounds for the Japanese *a*, right?

PC: [the differences between] cut, cat

Running words together:

PF: It's difficult for me. If this tool didn't listen and identify my speech, I spoke slower and more carefully. But, unfortunately, it became more difficult to run words together.

PA: I learn even if I pronounce two or more word together, TalkTyper can recognize them, so I got some hints for speech fluency.

PB: This software made me concentrated in how I pronounce English sentences and how to read fluently, how it flows.

Missing out sounds:

PF: I didn't notice this point in my pronunciation

PE: Sometimes it happened. I must be careful with the strength of my voice.

Strong/weak sounds, word stress:

PE: The same as before (missing out sounds)

PC: Polyphenols / po-LY-phe-nols?

PA: I tried to stress the stronger syllable of a word.

The first word is difficult for Talktyper to recognise, like *a* or *the*. And Talktyper misunderstands my voice, like *a dog* becoming *added*, or something like this.

PF: I sometimes checked [word stress] by using my dictionary.

Note. Because of participants' varying levels of proficiency, some of the spoken text is not quoted verbatim. PA = Participant A, PB = Participant B, etc. When there was more than one response from a participant for a given topic, each response is written as a new paragraph.