

A METHOD AND SYSTEM FOR ASSISTING IN IMPROVING SPEECH OF A USER IN A DESIGNATED LANGUAGE

Field of the Invention

- 5 The present invention relates to a method and system for assisting in improving speech of a user in a designated language; particularly, to receiving text in the designated language, or in another language, from the user to be spoken by the user in the designated language.
- 10 The present invention has particular but not exclusive application in comparing properties of the user speaking the text to corresponding expected properties of speech of the text to determine an accuracy of the user speaking the text and outputting feedback to the user indicative of their accuracy.

15 Background of the Invention

- Traditionally, persons wishing to learn a desired language will take a suitable course with a teacher and other students in the course. In the course, for example, the teacher provides the students with some text to be spoken and then provides feedback to the students as to how they spoke that text in comparison to the expected
- 20 speech of that text. The feedback may also include comments in relation to specific properties of speech of the text, such as fluency and pronunciation. Additionally, the teacher can provide exercises for the student to practice at home. In this case, however, the student does not receive any feedback at home and can, in some circumstances, develop bad habits detracting from the student's ability to learn and
- 25 improve speech of the desired language.

- In an existing example, electronic speech feedback systems are employed to remove the need for a physical teacher for the students to learn speech of a desired language. In the existing example, words, having known speech properties, such as
- 30 pronunciation, are presented to a user for the user to recite verbally. The feedback system receives audio of the user speaking the predetermined words and applies speech recognition algorithms to determine whether the user spoke those words accurately. In this example, however, users may find the predetermined, and often

repeated, words tedious and irrelevant; thus, potentially losing interest in improving their speech in the desired language.

Summary of the Invention

5 According to a first aspect of the present invention, there is provided a method of assisting in improving speech of a user in a designated language, the method including: receiving text in the designated language, or in another language, from the user to be spoken by the user in the designated language; processing the text to derive one or more expected properties of speech of the text in the designated
10 language; receiving audio of the user speaking the text in the designated language; processing the audio to derive one or more properties of the user speaking the text; comparing said one or more properties of the user speaking the text to corresponding ones of said one or more expected properties of said speech of the text to determine an accuracy of the user speaking the text; and outputting feedback to the user
15 indicative of the accuracy.

In an embodiment, the one or more expected properties of speech of the text and the one or more properties of the user speaking the text include at least one of pronunciation, fluency, and prosodic features. The prosodic features of speech
20 include variation in syllable length, loudness and pitch. It will be appreciated by those persons skilled in the art that the expected pronunciation, fluency, and prosodic features can be determined by empirical analysis of speech of speakers of the designated language. Indeed, the prosodic features can be indicative of accents of the designated language. In this case, for instance, the prosodic features of say a
25 Midwestern American accent is determined by empirical analysis and the method can determine and output feedback to the user indicative of the accuracy of the user speaking in the Midwestern American accent.

It will be appreciated by those persons skilled in the art that determining the accuracy
30 of the user speaking the text includes determining a measure of quality (e.g. 1/100 to 100/100) of the user speaking the text based on a sum of determinations of accuracy on each of the properties of the user speaking the text with corresponding ones of the expected properties of speech of the text. Accordingly, in the embodiment, the method determines the accuracy of the user speaking the text by determining the

accuracy of each of designated properties, such as pace, pitch, energy, pronunciation, fluency, etc. of the user speaking the text.

In an embodiment, the method further includes parsing the text into one or more sentences or segments of words. For example, a collection of words is parsed into one or more sentences by sequentially searching the text including those words for occurrences of unquoted sentence termination punctuation marks, such as periods, question marks, and exclamation marks.

Alternatively, the collection of words is parsed into a set of fixed word count segments typically two to eight words in size. The segments are then outputted and thus taught to the user sequentially, initially teaching one segment and then adding the second segment, and so forth until the entire collection of words is taught. In one variation, the segments can be built left-to-right from the start of the word collection until the entire collection is taught. Alternatively, the segments can be built right-to-left from the end of the word collection towards its being; for example, if the segment word count is three, initially the last three words of the collection would be taught then the last six words, then the last nine words, etc., until the collection is completed.

In another embodiment, a grammar of the one or more sentences of the text is also generated. It will be appreciated by those persons skilled in the art that the grammar (e.g. grammars) is a speech recognition grammar; that is, a formal grammatical structure that can be recognised by a speech recogniser implementing speech recognition algorithms. In an example, a prompt for the user to speak a first one of the sentences is provided to the user, such as via text of the first one of the sentences being displayed on a display of a user device. The method further includes receiving audio of the user speaking the first one of the sentences of the text then outputting feedback to the user indicative of the accuracy of the user speaking the first one of the sentences. After outputting the feedback, the user is then prompted to speak a second one of the sentences.

Thus, in an example of use, the user first inputs text in the designated language that they wish to learn to speak, or in another language and it is translated into the designated language for display to the user. The text is then parsed into sentences.

The user, wishing to improve their speech in the designated language, is then prompted to speak the displayed first sentence of the text. The audio of the user speaking the sentence is received and processed by the speech recognizer so that the user can receive feedback on their pronunciation, fluency, etc. of the sentence.

- 5 After the feedback is received, the user can then go on to the next sentence and receive feedback accordingly, and so on.

For example, the method uses Speech Recognition Grammar Specification (SRGS), for controlling all recognition operation together with a speech recognizer

- 10 implementing speech recognition algorithms. SRGS is a World Wide Web Consortium (W3C) standard for speech recognition grammars. It will also be appreciated that a speech recognition grammar is a set of word patterns instructing the speech recogniser as to what to expect a human to say.

- 15 In an embodiment, the method further includes processing the text to render the text as audio in the designated language and then outputting the audio to the user via a speaker of the user's device. That is, in this embodiment, a text to speech generator is employed to assist the user in improving speech of the designated language. For example, in use, the user speaks a sentence, receives feedback as to the accuracy of
- 20 the way the user spoke the sentence and then hears the expected speech of the sentence for comparison. Indeed, the sequence of using the method could be, for example, (a): input text, input speech then provide the expected speech output or (b): input text, provide expected speech output, then the speech input for comparison.

- 25 It will be appreciated by those persons skilled in the art that the step of processing the audio to derive the one or more properties of the user speaking the text is implemented using speech recognition algorithms and the step of processing the text to derive the one or more expected properties of speech of the text is implemented using speech synthesis algorithms.

30

Preferably, the accuracy includes more than one confidence value associated with the accuracy of the one or more properties of the user speaking the text corresponding to the one or more expected properties of speech of the text. The confidence value could, for example, be a PASS/FAIL arrangement based on how close the user's

spoken properties are to the expected properties. For instance, a baseline of a 70% match between the spoken fluency and pronunciation and the expected fluency and pronunciation of someone speaking the text is set and a PASS value is determined if the spoken fluency and pronunciation of the text exceeds the 70% match. In an example, the confidence value associated with the accuracy includes: HIGHLY ACCURATE (e.g. 80%+ match), marginally accurate (e.g. 50% - 80% match), marginally poor (e.g. 20% - 50% match), and very poor (e.g. 0% - 20% match). With reference to this example, the feedback also includes colours indicative of different ones of the confidence values of the accuracy. For instance, green is displayed on a display of a user device for HIGHLY ACCURATE, orange for marginally accurate, brown for marginally poor, and red for very poor. It is envisaged that audio feedback can also be provided to the user, such as outputting the speech "VERY POOR" via a speaker of the user device. Other forms of feedback include numerical grade (e.g. 1-10), letter grade (e.g. A – F), badges, or some other visual indicators of feedback.

In another embodiment, the method further includes processing the text to derive predetermined prohibited words in the text so that the prohibited words can be censored. For example, swear words and words indicative of hate speech, etc. are predetermined and stored in a memory so that they are accessible before implementing the step of parsing the text into sentences.

In an embodiment, the method further includes recording the audio of the user speaking the text in the designated language in a memory. In this way, the method can, say, process instances of the audio recorded in the database to derive the one or more expected properties of speech of the text. That is, for example, the expected pronunciation of text can be determined by analysis of recordings of many users speaking that text. In addition, the recorded speech of a particular user can be used for later analysis, such as indicating progress of that user speaking the designated language. In a further example, all text entered by users and received is retained for off-line analysis. That is, the collection of text strings are searched for repetitions (e.g. same input from different users) and for particular words or themes which are popular across multiple users.

According to another aspect of the present invention, there is provided a system for assisting in improving speech of a user in a designated language, the system including a processor having: an input module configured to: receive text in the designated language, or in another language, from the user to be spoken by the user
5 in the designated language; and receive audio of the user speaking the text in the designated language; a processing module configured to: process the text to derive one or more expected properties of speech of the text in the designated language; process the audio to derive one or more properties of the user speaking the text; and compare said one or more properties of the user speaking the text to corresponding
10 ones of said one or more expected properties of said speech of the text to determine an accuracy of the user speaking the text; and an output module configured to output feedback to the user indicative of the accuracy.

According to another aspect of the present invention, there is provided a system for
15 assisting in improving speech of a user in a designated language, the system including: a display configured to display text in the designated language, or in another language; a text input means configured to input said text in the designated language, or in said another language by the user to be spoken by the user in the designated language; a microphone configured to input audio of the user speaking
20 the text in the designated language; and a processor having: an input module configured to: receive the text inputted in said designated language, or in said another language; and receive the audio of the user speaking the text in the designated language; a processing module configured to: process the text to derive one or more expected properties of speech of the text in the designated language; process the
25 audio to derive one or more properties of the user speaking the text; and compare said one or more properties of the user speaking the text to corresponding ones of said one or more expected properties of said speech of the text to determine an accuracy of the user speaking the text; and an output module configured to output feedback to the user indicative of the accuracy.

30

In an embodiment, the system further includes a server including the above processor in data communication over a network with a user device including the display, the text input means and the microphone. That is, in this embodiment, the user has a user device having input and output functionality (e.g. a tablet, a personal computer,

or a smartphone) that is in data communication over a network (e.g. the Internet) with a server hosting the processor. Accordingly, the user inputs text to be spoken in the designated language by, say, typing the text or electronically pasting the text from a document into an allocated text box. The text is communicated over the network to
5 the server which parses the text into sentences and outputs the first sentence to be spoken by the user over the network to be displayed on the display of the user device. The user can then speak the first sentence and audio of the user speaking is captured by the microphone and communicated to the server for processing so that feedback can be determined and outputted via, say, the display of the user device to the user
10 as described above. That is, the display of the user device is further configured to display the feedback to the user.

In another embodiment, the user device including the display, text input means and the microphone also includes the processor. Thus, in this embodiment, the
15 processing is performed locally on the user device.

Preferably, the processing module is further configured to parse the text into one or more sentences and the output module is further configured to output a prompt for the user to speak a first one of the sentences. Also, the input module is further
20 configured to receive audio of the user speaking the first one of the sentences of the text and the output module is further configured to output feedback to the user indicative of the accuracy of the one or more properties of the user speaking the first one of the sentences of the text corresponding to the one or more expected properties of speech of the first one of the sentences of the text.

25 As described in some of the embodiments above, the system allows user generated text to be used in assisting the user in improving their speech in a designated language. The text is parsed into sentences and the user is prompted to speak one sentence at a time so that feedback can be received and considered by the user at
30 the conclusion of each sentence. Accordingly, the output module is further configured in this case to output a further prompt for the user to speak a second one of the sentences after outputting the feedback to the user indicative of the accuracy of the one or more properties of the user speaking the first one of the sentences of the text.

Also, the processing module is further configured to generate a grammar of the one or more sentences of the text and the prompt includes text of the first one of the sentences.

- 5 In an embodiment, the processing module is further configured to process the text to derive an expected speech of the text in the designated language. In the embodiment the output module is configured to output the expected speech of the text to be outputted to the user via a speaker of a user device. For example, the processing module includes a speech synthesis module to process the text to derive an expected
10 speech which is outputted to the user, for example after they speak their sentence, so that the user can compare and further improve their speech.

- In an embodiment, the processing module is further configured to process the text to derive predetermined prohibited words in the text so that the prohibited words can be
15 censored as described. The prohibited words are stored in a memory which can be located remote from the processor and accessible over a network or can be located locally.

- According to another aspect of the present invention, there is provided computer
20 programme code which when executed implements the above method.

According to another aspect of the present invention, there is provided a tangible computer readable medium including the above programme code.

- 25 According to another aspect of the present invention, there is provided a data file including the above programme code.

Brief Description of the Drawings

- In order that the invention can be more clearly understood, examples of embodiments
30 will now be described with reference to the accompanying drawings, wherein:

Figure 1 is a flow chart of a method of assisting in improving speech of a user in a designated language according to an embodiment of the present invention;

Figure 2 is a schematic view of a system for assisting in improving speech of a user in a designated language according to an embodiment of the present invention; and

- 5 Figure 3 is a further schematic view of the system of Figure 2 showing the system in communication with user device over a network.

Detailed Description

According to an embodiment of the present invention there is provided a method 10 of
10 assisting in improving speech of a user in a designated language, as shown in Figure 1. The method 10 includes the steps of receiving 12 text in the designated language, or in another language, from the user to be spoken by the user in the designated language, processing 14 the text to derive one or more expected properties of speech of the text in the designated language, receiving 16 audio of the user speaking the
15 text in the designated language, processing 18 the audio to derive one or more properties of the user speaking the text, comparing 20 one or more properties of the user speaking the text to corresponding ones of one or more expected properties of speech of the text to determine an accuracy of the user speaking the text, and outputting 22 feedback to the user indicative of the accuracy.

20 As described, the one or more expected properties of speech of the text and properties of the user speaking the text include at least one of pronunciation, fluency, and prosodic features. The prosodic features of speech include variation in syllable length of words, loudness and pitch. Thus, in an example of use, fluency and, say,
25 pronunciation of expected speech of the inputted text are derived from the text and compared against the fluency and pronunciation derived from the received audio of the user speaking the text to determine an accuracy of the user speaking the text. The accuracy of the user speaking, in terms of their pronunciation and fluency, is then fed back to the user to assist them in improving their speech in the designated
30 language (e.g. English).

In another embodiment of the present invention, there is provided a system 24, as shown in Figure 2, for assisting in improving speech of a user in a designated

language that implements the method 10. The system 24 includes a processor 26 having a number of a number of modules for implementing the method 10. Namely, the processor 26 includes an input module 28, processing module 30 and an output module 32. The processor 26 is arranged to receive and transmit information over, say, a network and/or between other components of the system 24, such as a memory 52 (shown in Figure 3) via a communication channel 34. In the embodiment, the processor 26 is implemented by a computer in communication over the communication channel 34 with input devices and output devices contained in a user device 38 (shown in Figure 3). As described, however, it is envisaged that the memory 52 could also reside on, say, a server remote from the processor 26 which is accessible over a network. In any event, it will be appreciated by those persons skilled in the art that the input 28 and output 32 modules have suitable interfaces for interfacing with the network, the modules in the system 24, and establishing the communications channel 34. Furthermore, it will also be appreciated that the input devices and the output devices need not be contained in the same user device 38. For example, text to be spoken by a user can be displayed on a television in communication with the system 24 over a network whilst audio is recorded by the user using a smartphone in communication with the system 24 over the network.

In any case, as described, the input module 28 is configured to receive text in the designated language, or in another language, from the user and to receive audio of the user speaking the text in the designated language. The processing module 30 is configured to process the text to derive expected properties (e.g. fluency) of speech of the received text from the input module 28, using, for instance, speech recognition algorithms, and to process the received audio from the input module 28 to derive properties of the user speaking the text. The processing module 30 is also used to parse the text into one or more sentences. That is, the received text, including a collection of words, is parsed into sentences by the processing module 30 sequentially searching the text for occurrences of unquoted sentence termination punctuation marks, such as periods, question marks, and exclamation marks. Thus, in use, the user is prompted to speak a first one of the parsed sentences and audio of the user speaking the first one of the sentences of the text is received by the input module 28 and so on.

In an example, the sentence: *He said, "I am not done!" then he continued speaking.* is parsed into a single sentence: *He said, "I am not done!" then he continued speaking.* However, the following sentence: *He said, I am not done! Then he continued speaking.* is parsed into two sentences: 1. *He said, I am not done!*; 2. *Then he continued speaking.* for prompting of the user to speak those sentences. Further, once the collection of words from the inputted text has been parsed into individual sentences, each sentence is converted to a speech recognition grammar consistent with its expected properties. The sentence is presented visually or acoustically to the user, then the user is prompted to speak what he or she has read or heard, and the user's verbal response is passed from the input module 28 to a speech recogniser implemented by the processing module 30, with the sentence grammar, for analysis.

Accordingly, the processing module 30 compares the properties of the user speaking the text to the expected properties of speech of the text, which is for instance imposed by a speech recognition vendor, to determine an accuracy of the user speaking the text in relation to the properties. The output module 32 receives the determination of accuracy from the processing module 30 and outputs feedback to the user indicative of the accuracy of the user speaking the text.

Figure 3 shows a system 36 for assisting in improving speech of a user in a designated language including a user device 38 for inputting and outputting information to the processor 26 over a network 40, such as the Internet. As described, it is envisaged that, in another embodiment not shown in the figures, the system 24 is self-contained and includes suitable components to input and output information to the processor 26 to implement the method 10. Nonetheless, the processor 26 is shown as being implemented by a server 54 that is accessible over the Internet 40. It is also envisaged that the processor 26 and the memory 52 can be implemented as, say, a cloud service with virtual servers or across multiple physical servers. In one example, the server 54 is accessible via a Uniform Resource Locator (URL), which can be embedded in websites for users to access via their user devices 38. In another example, the network 40 is a Local Area Network (LAN) and the user device 38 communicates with the server via, say, WiFi.

The user device 38 shown in Figure 3 includes a display 44 configured to display text in the designated language, or in another language, to the user. It can be seen in the example exemplified in Figure 3 that the text displayed is "Hey Joe, what are you doing" and the user device is a tablet computer with touch screen or gesture-reading capabilities. As described, other user devices are envisaged for use with the system 36, such as smartphones and personal computers.

With reference to an example using this embodiment, the user has inputted the text: "Hey Joe, what are you doing" using an input means 46 taking the form of a touch screen keyboard as part of the text to be spoken by the user to assist the user in learning English. Once the user is finished typing in the text, or pasting the text from another document, the user can then hit a "submit" button (not shown) and transmit the text over the Internet 40 to the server 54. The inputted text is transmitted over communication channels 42 and 34, via the Internet 40, to the input module 28 so that the processing module 30 can process the text to derive the expected properties of speech of the text. As described, the processing module 30 parses the text into sentences to be displayed to the user which are to be spoken one sentence at a time so as to receive feedback indicative of the accuracy of how the text was spoken at the conclusion of each sentence. Furthermore, in an embodiment, the processing module 30 converts each sentence into context free grammar, which is the syntax required by the target speech recogniser employed by the processing module 30 for analyzing the speech of the user speaking the sentences. That is, in this embodiment, the grammar represents the structure of text packaged to be conveyed to the speech recogniser.

25

With reference to the same example, the text "Hey Joe, what are you doing" is outputted back to the user device 38, via the output module 32, and displayed on the display 44 as a prompt for the first sentence to be spoken. That is, the prompt includes text of the first sentence to be spoken. A microphone 48 is used to record the user speaking this sentence and the audio of the user speaking the sentence is transmitted, via the Internet 40, to the input module 28 so that the processing module 30 can process the audio to derive properties of the user speaking the text. The processing module 30 then compares the derived properties of the audio with the

30

expected properties to determine the accuracy of the user speaking the text using the target speech recogniser algorithms.

The output module 32 then outputs feedback to the user indicative of the accuracy determined by the processing module 30, via the Internet 40, as colours to be displayed on the display 44 indicative of different levels of confidence values of the determined accuracy. As described, the colour green is used to indicate a confidence value associated with the accuracy being HIGHLY ACCURATE (e.g. 80%+ match). In another example, the output module 32 outputs feedback to the user indicative of the accuracy determined by the processing module 30 as voice outputted from speaker 50 of the user device 38. For example, the voice feedback states "Highly accurate" over the speaker 50 when a confidence value of greater than 80% is determined by the processing module 30. As described the display 44 can also be configured to display other protocols indicative of different levels of confidence values of the determined accuracy, such as numerical and letter grades. Furthermore, in an embodiment, the processing module 30 includes speech synthesis algorithms, as described, and here the output module 32 can be configured to output synthesised expected speech of the text to be outputted from the speaker 50 to the user to further assist the user in improving their speech.

After outputting the feedback, the output module 32 is further configured to output the next sentence (not shown) parsed by the processing module 30 from the user inputted text as a further prompt for the user to speak the second sentence. The prompt including the text of the second sentence is also displayed on the display 44 and the microphone 48 is used to record the user speaking this sentence. As with the first sentence, the audio of the user speaking this sentence is transmitted to the input module 28 so that the processing module 30 can process the audio to derive properties of the user speaking the text. The processing module 30 then compares the derived properties of the audio with the expected properties to determine the accuracy of the user speaking the text and the output module 32 outputs feedback to the user indicative of the accuracy determined by the processing module 30 as colours displayed on the display 44. The output module then outputs the next sentence and repeats the process until all sentences parsed from the user inputted text are spoken or the user terminates the process prematurely.

As described, the memory 52 can be used to record the audio of the user speaking the text in the designated language. Accordingly, the processing module 30 can further process instances of the audio recorded in the memory 52 to derive and refine the one or more expected properties of speech of the text as the expected pronunciation of text can be determined from analysis of recordings of users speaking that text. In addition, the recorded speech is stored in the memory 52 in association with data indicative of particular users so that the recordings can be retrieved for later analysis, such as for indicating progress of a particular user in their speech of, say, English. It will also be appreciated by those persons skilled in the art that the method and system can be employed for users wishing to improve their speech in other languages, such as French, Chinese, Japanese, etc.

Further aspects of the method will be apparent from the above description of the system. A person skilled in the art will also appreciate that the method could be embodied in a program code. The program code could be supplied in a number of ways, for example on tangible computer readable medium, such as a disk or memory or as a data signal or data file.

It is to be understood to a person skilled in the art of the invention that many alterations, additions and/or modifications may be made without departing from the spirit and scope of the invention.

It is to be understood that, if any prior art is referred to herein, such reference does not constitute an admission that the prior art forms a part of the common general knowledge in the art in any country.

The present invention may be used as the basis for priority in respect of one or more future applications, and the claims of such future applications may be directed to any one feature or combination of features that are described in the present application. As such, future applications include one or more of the following claims, which are given by way of example and are non-limiting with regard to what may be claimed in any future application.

Claims:

1. A method of assisting in improving speech of a user in a designated language, the method including:

5 receiving text in the designated language, or in another language, from the user to be spoken by the user in the designated language;
processing the text to derive one or more expected properties of speech of the text in the designated language;

receiving audio of the user speaking the text in the designated language;
10 processing the audio to derive one or more properties of the user speaking the text;

comparing said one or more properties of the user speaking the text to corresponding ones of said one or more expected properties of said speech of the text to determine an accuracy of the user speaking the text; and

15 outputting feedback to the user indicative of the accuracy of the user speaking the text, wherein the method further includes:

parsing the text into one or more sentences or segments of words to be spoken by the user;

prompting the user to speak one of the sentences or segments of words at a
20 time;

receiving audio of the user speaking the sentences or segments of the text;
and

outputting said feedback to the user indicative of the accuracy of the user speaking the sentences or segments of the text.

25

2. A method as claimed in claim 1, further including outputting a prompt for the user to speak a first one of the sentences or segments and receiving audio of the user speaking the first one of the sentences or segments of the text.

30 3. A method as claimed in claim 2, further including outputting said feedback to the user indicative of the accuracy of the user speaking the first one of the sentences or segments of the text.

4. A method as claimed in claim 3, further including outputting a further prompt for the user to speak a second one of the sentences or segments after outputting the feedback to the user indicative of the accuracy of the user speaking the first one of the sentences or segments of the text.

5

5. A method as claimed in claim 2, wherein the prompt includes text of the first one of the sentences or segments to be displayed on a display of a user device to the user.

10

6. A method as claimed in claim 1, further including generating a speech recognition grammar of the text.

15

7. A method as claimed in any one of claims 1 to 6, further including processing the text to derive a synthesised expected speech of the text in the designated language.

8. A method as claimed in claim 7, further including outputting the synthesised expected speech of the text to the user via a speaker of a user device.

20

9. A method as claimed in any one of claims 1 to 8, wherein the accuracy includes confidence values associated with the accuracy of the one or more properties of the user speaking the text corresponding to the one or more expected properties of speech of the text.

25

10. A method as claimed in claim 9, wherein the feedback includes colours indicative of different confidence values of the accuracy.

30

11. A method as claimed in any one of claims 1 to 10, further including processing the text to derive predetermined prohibited words in the text so that the prohibited words can be censored.

12. A method as claimed in any one of claims 1 to 11, wherein the one or more expected properties of speech of the text and the one or more properties of the user speaking the text include at least one of pronunciation, fluency, and prosodic features.

13. A method as claimed in any one of claims 1 to 12, further including processing the audio to derive said one or more properties of the user speaking the text using speech recognition algorithms.

5

14. A method as claimed in any one of claims 1 to 13, further including processing the text to derive said one or more expected properties of the text using speech synthesis algorithms.

10 15. A system for assisting in improving speech of a user in a designated language, the system including a processor having:

an input module configured to:

receive text in the designated language, or in another language, from the user to be spoken by the user in the designated language; and

15 receive audio of the user speaking the text in the designated language;

a processing module configured to:

process the text to derive one or more expected properties of speech of the text in the designated language;

20 process the audio to derive one or more properties of the user speaking the text; and

compare said one or more properties of the user speaking the text to corresponding ones of said one or more expected properties of said speech of the text to determine an accuracy of the user speaking the text; and

25 an output module configured to output feedback to the user indicative of the accuracy of the user speaking the text, wherein

the processing module is further configured to parse the text into one or more sentences or segments of words;

the output module is further configured to output prompts for the user to speak one of the sentences or segments of words at a time;

30 the input module is further configured to receive audio of the user speaking the one or more sentences or segments of the text; and

the output module is further configured to output said feedback to the user indicative of the accuracy of the user speaking the sentences or segments of the text.

16. A system as claimed in claim 15, wherein the output module is further configured to output a prompt for the user to speak a first one of the sentences or segments.

5

17. A system as claimed in claim 16, wherein the input module is further configured to receive audio of the user speaking the first one of the sentences or segments of the text.

10 18. A system as claimed in claim 17, wherein the output module is further configured to output said feedback to the user indicative of the accuracy of the user speaking the first one of the sentences or segments of the text.

15 19. A system as claimed in claim 18, wherein the output module is further configured to output a further prompt for the user to speak a second one of the sentences or segments after outputting the feedback to the user indicative of the accuracy of the user speaking the first one of the sentences or segments of the text.

20 20. A system as claimed in claims 16, wherein the prompt includes text of the first one of the sentences or segments to be displayed on a display of a user device to the user.

25 21. A system as claimed in claim 15, wherein the processing module is further configured to generate a speech recognition grammar of the text.

22. A system as claimed in any one of claims 15 to 21, wherein the processing module is further configured to process the text to derive a synthesised expected speech of the text in the designated language.

30 23. A system as claimed in claim 22, wherein the output module is further configured to output the synthesised expected speech of the text to the user via a speaker of a user device.

24 A system as claimed in any one of claims 15 to 23, wherein the accuracy includes confidence values associated with the accuracy of the one or more properties of the user speaking the text corresponding to the one or more expected properties of speech of the text.

5

25. A system as claimed in claim 24, wherein the feedback includes colours indicative of different confidence values of the accuracy.

26. A system as claimed in any one of claims 15 to 25, wherein the processing module is further configured to process the text to derive predetermined prohibited words in the text so that the prohibited words can be censored.

10

27. A system as claimed in any one of claims 15 to 26, wherein the system further includes a memory for recording the audio of the user speaking the text.

15

28. A system as claimed in claim 27, wherein the processing module is further configured to process instances of the audio recorded in the memory to derive the one or more expected properties of speech of the text.

29. A system as claimed in any one of claims 15 to 28, wherein the one or more expected properties of speech of the text and the one or more properties of the user speaking the text include at least one of pronunciation, fluency, and prosodic features.

20

30. A system for assisting in improving speech of a user in a designated language, the system including:

25

a display configured to display text in the designated language, or in another language;

a text input means configured to input said text in the designated language, or in said another language by the user to be spoken by the user in the designated

30

language;

a microphone configured to input audio of the user speaking the text in the designated language; and

a processor having:

an input module configured to:

receive the text inputted in said designated language, or in said another language; and

receive the audio of the user speaking the text in the designated language;

5 a processing module configured to:

process the text to derive one or more expected properties of speech of the text in the designated language;

process the audio to derive one or more properties of the user speaking the text; and

10 compare said one or more properties of the user speaking the text to corresponding ones of said one or more expected properties of said speech of the text to determine an accuracy of the user speaking the text; and

15 an output module configured to output feedback to the user indicative of the accuracy of the user speaking the text, wherein

the processing module is further configured to parse the text into one or more sentences or segments of words;

the output module is further configured to output prompts for the user to speak one of the sentences or segments of words at a time;

20 the input module is further configured to receive audio of the user speaking the one or more sentences or segments of the text; and

the output module is further configured to output said feedback to the user indicative of the accuracy of the user speaking the sentences or segments of the text.

25

31. A system as claimed in claim 30, wherein the display is further configured to display the feedback to the user.

30 32. A system as claimed in claim 31, further including a server including the processor in data communication over a network with a user device including the display, the text input means and the microphone.

33. Computer programme code which when executed implements the method of any one of claims 1 to 14.

34. A tangible computer readable medium including the programme code of claim 33.
- 5 35. A data file including the programme code of claim 33.