Claims

WHAT IS CLAIMED IS:

1. A text-to-speech (TTS) conversion method, comprising:

obtaining text which is to be segmented into character substrings, each character substring comprises multiple words, and in two adjacent character substrings, the word at the end of the preceding character substring is identical to the word at the beginning of the next character substring, said text is in Chinese and is to be segmented using Chinese word segmentation methods;

generating audio units corresponding to said character substrings, and in the audio units corresponding to two adjacent character substrings, similar audio frames are existing between the audio frames at the end of the preceding audio unit and the audio frames at the beginning of the next audio unit;

concatenating the beginning of an audio unit to the end of the preceding audio unit from the location where both audio units have similar audio frames, thus generating speech corresponding to the two adjacent character substrings.

2. The method described in Claim 1, wherein said text is segmented into at least one character substring comprising: based on a predefined length of character, the length of character substrings after segmentation is equal to the predefined length of character sequences.

3. The method described in Claim 1, wherein generating audio units corresponding to said character substrings comprises: inputting the character substrings into a neural network identification model which identifies audio units corresponding to the character substrings.

4. The method described in Claim 1, wherein based on a predetermined parallel threshold, character substrings, whose quantity is equivalent to the said parallel threshold value, are converted into corresponding audio units through parallel conversion.

5. A TTS conversion apparatus comprising:

a text segmentation module, which is for obtaining text to be segmented into character substrings, each character substring comprises multiple words, and in two adjacent character substrings, the word at the end of the preceding character substring is identical to the word at the beginning of the next character substring, said text is in Chinese and is to be segmented using Chinese word segmentation methods;

an audio conversion module, which is for generating audio units corresponding to said character substrings, and in the audio units corresponding to two adjacent character substrings, similar audio frames are existing between the audio frames at the end of the preceding audio unit and the audio frames at the beginning of the next audio unit;

an audio synthesis module, which is for concatenating the beginning of an audio unit to the end of the preceding audio unit from the location where both audio units have similar audio frames, thus generating speech corresponding to the two adjacent character substrings.

6. The apparatus described in Claim 5, wherein a text segmentation module segments said text into at least one character substring comprising: based on a predefined length of character, the length of character substrings after segmentation is equal to the predefined length of character sequences.

7. The apparatus described in Claim 5, wherein said audio conversion module generate audio units corresponding to said character substrings, the apparatus comprising: inputting, by the audio conversion module, the character substrings into a neural network identification model which identifies audio units corresponding to the character substrings.

8. The apparatus described in Claim 5, wherein based on a predetermined parallel threshold, character substrings, whose quantity is equivalent to the said parallel threshold value, are generated by said audio synthesis module and said audio conversion module into corresponding audio units through parallel processing.

9. A device, comprising:

a memory that stores instructions, and a processor which is for enabling said device to read and execute the instructions stored in said memory, thereby implementing any of the method as described in Claims 1 through 4.

10. A machine-readable medium, wherein stored instructions are executed by a machine that implements any of the method as described in Claims 1 through 4.