CLAIMS

1. A text information processing method, comprising:

acquiring (S201) a phoneme vector corresponding to an individual phoneme and a semantic vector corresponding to the individual phoneme in text information;

acquiring (S202) first semantic information output at a previous moment, wherein the first semantic information is semantic information corresponding to part of the text information in the text information, and the part of the text information is text information that has been converted into voice information;

determining (S203) a context vector corresponding to a current moment according to the first semantic information, the phoneme vector corresponding to the individual phoneme and the semantic vector corresponding to the individual phoneme; and

determining (S204) voice information at the current moment according to the context vector and the first semantic information;

wherein acquiring (S201) the phoneme vector corresponding to the individual phoneme in the text information comprises:

acquiring (S301, S501) a phoneme sequence of the text information, wherein the phoneme sequence includes at least one phoneme; and

determining the phoneme vector corresponding to the individual phoneme according to the at least one phoneme and a preset corresponding relationship between the phoneme and the phoneme vector;

wherein acquiring (S201) the semantic vector corresponding to the individual phoneme comprises:

encoding (S304, S504), by a preset encoder, the phoneme vector corresponding to the individual phoneme to obtain the semantic vector corresponding to the individual phoneme;

wherein determining (S204) the voice information at the current moment according to the context vector and the first semantic information comprises:

decoding, by a preset decoder, the context vector and the first semantic information to obtain second semantic information corresponding to the current moment; and

performing (S510) Mel spectrum network mapping processing on the second semantic information to obtain the voice information at the current moment.

2. The method according to claim 1, wherein determining (S203) the context vector corresponding to the current moment according to the first semantic information, the phoneme vector corresponding to the individual phoneme, and the semantic vector corresponding to the individual phoneme comprises:

determining a semantic matching degree according to the first semantic information and the semantic vector corresponding to the individual phoneme; and

determining the context vector corresponding to the current moment according to the semantic matching degree, the phoneme vector corresponding to the individual phoneme, and the semantic vector corresponding to the individual phoneme.

3. The method according to claim 2, wherein determining the semantic matching degree according to the first semantic information and the semantic vector corresponding to the individual phoneme comprises:

processing (S306) the first semantic information and the semantic vector corresponding to the individual phoneme by a first preset model to obtain the semantic matching degree, wherein the first preset model is obtained by learning multiple groups of first samples, and each group of first samples includes semantic information and semantic vectors.

4. The method according to claim 3, wherein determining the context vector corresponding to the current moment according to the semantic matching degree, the phoneme vector corresponding to the individual phoneme and the semantic vector corresponding to the individual phoneme comprises:

determining (S307) a semantic context vector according to the semantic matching degree and the semantic vector corresponding to the individual phoneme;

determining (S308) a phoneme context vector according to the semantic matching degree and the phoneme vector corresponding to the individual phoneme; and

combining (S309) the semantic context vector and the phoneme context vector to determine the context vector corresponding to the current moment.

5. The method according to claim 1, wherein determining (S203) the context vector corresponding to the current moment according to the first semantic information, the

phoneme vector corresponding to the individual phoneme and the semantic vector corresponding to the individual phoneme comprises:

determining (S506) a resultant vector corresponding to the individual phoneme according to the phoneme vector corresponding to the individual phoneme and the semantic vector corresponding to the individual phoneme;

determining the semantic matching degree according to the first semantic information and the resultant vector corresponding to the individual phoneme; and

determining (S508) the context vector corresponding to the current moment according to the semantic matching degree and the resultant vector corresponding to the individual phoneme.

6. The method according to claim 5, wherein determining the semantic matching degree according to the first semantic information and the resultant vector corresponding to the individual phoneme comprises:

processing (S507) the first semantic information and the resultant vector corresponding to the individual phoneme by a second preset model to obtain the semantic matching degree, wherein the second preset model is obtained by learning multiple groups of second samples, and each group of second samples includes semantic information and resultant vectors.

7. The method according to any one of claims 1 to 6, wherein determining the phoneme vector corresponding to the individual phoneme according to the at least one phoneme comprises:

finding (S302, S502) an identifier corresponding to the individual phoneme in a preset phoneme list; and

searching (S303, S503) for the phoneme vector corresponding to the individual phoneme in a preset phoneme lookup table according to the identifier corresponding to the individual phoneme.

8. A text information processing apparatus (10), comprising: a first acquiring module (11), a second acquiring module (12), a first determining module (13) and a second determining module (14), wherein,

the first acquiring module (11) is configured to acquire a phoneme vector corresponding to an individual phoneme and a semantic vector corresponding to the individual phoneme in text information;

the second acquiring module (12) is configured to acquire first semantic information output at a previous moment, wherein the first semantic information is semantic information corresponding to part of the text information in the text information, and the part of the text information is text information that has been converted into voice information;

the first determining module (13) is configured to determine a context vector corresponding to a current moment according to the first semantic information, the phoneme vector corresponding to the individual phoneme and the semantic vector corresponding to the individual phoneme;

the second determining module (14) is configured to determine voice information at the current moment according to the context vector and the first semantic information;

wherein the first acquiring module (11) is specifically configured to:

acquire a phoneme sequence of the text information, wherein the phoneme sequence includes at least one phoneme; and

determine the phoneme vector corresponding to the individual phoneme according to the at least one phoneme and a preset corresponding relationship between the phoneme and the phoneme vector;

encode the phoneme vector corresponding to the individual phoneme to obtain the semantic vector corresponding to the individual phoneme;

wherein the second determining module (14) is specifically configured to:

decode the context vector and the first semantic information to obtain second semantic information corresponding to the current moment; and

perform Mel spectrum network mapping processing on the second semantic information to obtain the voice information at the current moment.

9. A text information processing apparatus, comprising: at least one processor and a memory;

the memory stores computer-executable instructions;

the at least one processor executes the computer-executable instructions stored in the memory, causing the at least one processor to execute the text information processing method of any one of claims 1 to 7.

10. A computer-readable storage medium, wherein computer-executable instructions are stored in the computer-readable storage medium and when a processor executes the computer-executable instructions, the text information processing method of any one of claims 1 to 7 is implemented.

11. A computer program product comprising a computer program, wherein when the computer program is executed by a processor, the text information processing method of any one of claims 1 to 7 is implemented.

12. A chip, comprising: a memory and a processor;

the memory stores a computer program;

the processor runs the computer program to execute the text information processing method of any one of claims 1 to 7.