WHAT IS CLAIMED IS:

1. A method for end-to-end speech enhancement based on a neural network, comprising:

obtaining a time-domain smoothing feature of an original speech signal by performing feature extraction on the original speech signal using a time-domain convolution kernel, wherein the time-domain convolution kernel is implemented in a Time-Domain Recursive Averaging Layer (TRAL) module, the time-domain smoothing feature of the original speech signal represents phase information and amplitude information in the original speech signal; and

obtaining an enhanced speech signal by performing combined feature extraction on the original speech signal and the time-domain smoothing feature of the original speech signal.

2. The method for end-to-end speech enhancement according to claim 1, wherein obtaining a time-domain smoothing feature of an original speech signal by performing feature extraction on the original speech signal using a time-domain convolution kernel comprises:

determining a time-domain smoothing parameter matrix according to a convolution sliding window and a time-domain smoothing factor;

obtaining a weight matrix of the time-domain convolution kernel by performing a product operation on the time-domain smoothing parameter matrix; and

obtaining the time-domain smoothing feature of the original speech signal by performing a convolution operation on the weight matrix of the time-domain convolution kernel and the original speech signal.

3. The method for end-to-end speech enhancement according to claim 2, wherein determining a time-domain smoothing parameter matrix according to a convolution sliding window and a time-domain smoothing factor comprises:

initializing a plurality of time-domain smoothing factors; and

obtaining the time-domain smoothing parameter matrix based on a preset convolution sliding window and the plurality of time-domain smoothing factors.

4. The method for end-to-end speech enhancement according to claim 1, wherein obtaining an enhanced speech signal by performing combined feature extraction on the original speech signal and the time-domain smoothing feature of the original speech signal comprises:

obtaining a speech signal to be enhanced by combining the original speech signal and the time-domain smoothing feature of the original speech signal;

training a weight matrix of the time-domain convolution kernel by using a back propagation algorithm with the speech signal to be enhanced as an input of a deep neural network, and;

obtaining the enhanced speech signal by performing combined feature extraction on the speech signal to be enhanced according to the weight matrix obtained by training.

5. The method for end-to-end speech enhancement according to claim 4, wherein training a weight matrix of the time-domain convolution kernel by using a back propagation algorithm with the speech signal to be enhanced as an input of a deep neural network comprises:

inputting the speech signal to be enhanced into the deep neural network, and constructing a time-domain loss function; and

training the weight matrix of the time-domain convolution kernel by using an error back propagation algorithm according to the time-domain loss function.

6. The method for end-to-end speech enhancement according to claim 4, wherein obtaining the enhanced speech signal by performing combined feature extraction on the speech signal to be enhanced according to the weight matrix obtained by training comprises:

obtaining a first time-domain feature map by performing a convolution operation on the weight matrix obtained by training and an original speech signal in the speech signal to be enhanced;

obtaining a second time-domain feature map by performing a convolution operation on the weight matrix obtained by training and a smoothing feature in the speech signal to be enhanced; and

obtaining the enhanced speech signal by combining the first time-domain feature map and the second time-domain feature map.

7. The method for end-to-end speech enhancement according to claim 1, wherein obtaining a time-domain smoothing feature of an original speech signal by performing feature extraction on the original speech signal using a time-domain convolution kernel comprises:

performing speech enhancement on the phase information and the amplitude information in the original speech signal by inputting the original speech signal into a deep neural network for time-varying feature extraction.

8. The method for end-to-end speech enhancement according to claim 1, wherein the

original speech signal is represented by a one-dimensional vector.

9. An apparatus for end-to-end speech enhancement based on a neural network, comprising:

a time-domain smoothing feature extraction module, configured to obtain a time-domain smoothing feature of an original speech signal by performing feature extraction on the original speech signal using a time-domain convolution kernel, wherein the time-domain convolution kernel is implemented in a Time-Domain Recursive Averaging Layer (TRAL) module, the time-domain smoothing feature of the original speech signal represents phase information and amplitude information in the original speech signal; and

a combined feature extraction module, configured to obtain an enhanced speech signal by performing combined feature extraction on the original speech signal and the time-domain smoothing feature of the original speech signal.

10. The apparatus for end-to-end speech enhancement according to claim 9, wherein the time-domain smoothing feature extraction module comprises:

a parameter matrix determination unit, configured to determine a time-domain smoothing parameter matrix according to a convolution sliding window and a time-domain smoothing factor;

a weight matrix determination unit, configured to obtain a weight matrix of the time-domain convolution kernel by performing a product operation on the time-domain smoothing parameter matrix; and

a time-domain operation unit, configured to obtain the time-domain smoothing feature of the original speech signal by performing a convolution operation on the weight matrix of the time-domain convolution kernel and the original speech signal.

11. The apparatus for end-to-end speech enhancement according to claim 10, wherein the parameter matrix determination unit comprises:

a data initialization subunit, configured to initialize a plurality of time-domain smoothing factors; and

a matrix determination subunit, configured to obtain the time-domain smoothing parameter matrix based on a preset convolution sliding window and the plurality of time-domain smoothing factors.

12. The apparatus for end-to-end speech enhancement according to claim 9, wherein the combined feature extraction module comprises:

an input signal obtaining unit, configured to obtain a speech signal to be enhanced by combining the original speech signal and the time-domain smoothing feature of the original speech signal;

a weight matrix training unit, configured to train a weight matrix of the time-domain convolution kernel by using a back propagation algorithm with the speech signal to be enhanced as an input of a deep neural network; and

an enhanced speech signal obtaining unit, configured to obtain the enhanced speech signal by performing combined feature extraction on the speech signal to be enhanced according to the weight matrix obtained by training.

13. The apparatus for end-to-end speech enhancement according to claim 12, wherein the weight matrix training unit comprises:

a data input subunit, configured to input the speech signal to be enhanced into a deep neural network, and construct a time-domain loss function; and

a data training subunit, configured to train the weight matrix of the time-domain convolution kernel by using an error back propagation algorithm according to the time-domain loss function.

14. The apparatus for end-to-end speech enhancement according to claim 12, wherein the enhanced speech signal obtaining unit comprises:

a first feature map obtaining subunit, configured to obtain a first time-domain feature map by performing a convolution operation on the weight matrix obtained by training and an original speech signal in the speech signal to be enhanced;

a second feature map obtaining subunit, configured to obtain a second time-domain feature map by performing a convolution operation on the weight matrix obtained by training and a smoothing feature in the speech signal to be enhanced; and

a feature combination subunit, configured to obtain the enhanced speech signal by combining the first time-domain feature map and the second time-domain feature map.

15. A computer-readable storage medium, with a computer program stored thereon; wherein, when the computer program is executed by a processor, the method according to any one of claims 1-8 is implemented.

16. An electronic device, comprising:

a processor; and

a memory, configured to store an executable instruction of the processor;

wherein the processor is configured to execute the method according to any one of claims 1-8 by executing the executable instruction.