device and a second device connected with each other via an audio interface, wherein

the first device is configured for modulating an identifier of a modulation mode supported by the first device and an identifier of a modulation mode supported by a second device using the modulation mode supported by the first device according to a predetermined rule to generate a first audio detection frame, sending the first audio detection frame to the second device, determining whether a first audio detection feedback frame is correct after receiving the first audio detection feedback frame from the second device, and if yes, demodulating an audio data frame from the second device using a demodulation mode corresponding to the modulation mode supported by the second device;

the second device is configured for determining whether the first audio detection frame is correct after receiving the first audio detection frame, and if yes, selecting the modulation mode supported by the second device corresponding to the identifier of the modulation mode supported by the second device from pre-stored modulation modes supported by the second device according to the identifier of the modulation mode supported by the second device carried in the first audio detection frame, and modulating the identifier of the modulation mode supported by the first device and the identifier of the modulation mode supported by the second device carried in the first audio detection frame using the modulation mode supported by the second device to generate the first audio detection feedback frame, sending the first audio detection feedback frame to the first device, and if the first device determines that the first audio detection feedback frame is correct, demodulating an audio data frame from the first device using a demodulation mode corresponding to the modulation mode supported by the first device.

9. The system according to claim 8, wherein the predetermined rule comprises combining a first modulation mode ~~and~~ a second modulation mode supported by the first device with a first modulation mode ~~and~~ a second modulation mode supported by the second device to obtain sets, setting a weight value for each set, and sequentially selecting the sets according to a magnitude of the weight values.

10. The system according to claim 9, wherein the first device is further configured for:

if the first device does not receive the first audio detection feedback frame, or if the first audio detection feedback frame is wrong,

modulating the identifier of the first modulation mode supported by the first device and an identifier of the second modulation mode supported by the second device using the first

a first determining module configured for determining whether a first audio detection feedback frame from the electronic signature token is received;

a second determining module configured for determining whether the first audio detection feedback frame is correct if the first determining module determines that the first audio detection feedback frame from the electronic signature token is received; and

a demodulating module configured for demodulating an audio data frame from the electronic signature token using a demodulation mode corresponding to the modulation mode supported by the electronic signature token, if the second determining module determines that the first audio detection feedback frame is correct.

16. The apparatus according to claim 15, wherein the predetermined rule comprises combining a first modulation mode ~~and~~ a second modulation mode supported by the apparatus with a first modulation mode ~~and~~ a second modulation mode supported by the electronic signature token to obtain sets, setting a weight value for each set, and sequentially selecting the sets according to a magnitude of the weight values.

17. The apparatus according to claim 16, wherein the modulating module is further configured for:

if the first determining module determines that the first audio detection feedback frame is not received, or if the second determining module determines that the first audio detection feedback frame is wrong,

modulating the identifier of the first modulation mode supported by the apparatus and an identifier of the second modulation mode supported by the electronic signature token using the first modulation mode supported by the apparatus according to the predetermined rule to generate a second audio detection frame, or

modulating an identifier of the second modulation mode supported by the apparatus and the identifier of the first modulation mode supported by the electronic signature token using the second modulation mode supported by the apparatus according to the predetermined rule to generate a third audio detection frame, or

modulating the identifier of the second modulation mode supported by the apparatus and the identifier of the second modulation mode supported by the electronic signature token using the second modulation mode supported by the apparatus according to the predetermined rule to generate a fourth audio detection frame,

device and a second device connected with each other via an audio interface, wherein

the first device is configured for modulating an identifier of a modulation mode supported by the first device and an identifier of a modulation mode supported by a second device using the modulation mode supported by the first device according to a predetermined rule to generate a first audio detection frame, sending the first audio detection frame to the second device, determining whether a first audio detection feedback frame is correct after receiving the first audio detection feedback frame from the second device, and if yes, demodulating an audio data frame from the second device using a demodulation mode corresponding to the modulation mode supported by the second device;

the second device is configured for determining whether the first audio detection frame is correct after receiving the first audio detection frame, and if yes, selecting the modulation mode supported by the second device corresponding to the identifier of the modulation mode supported by the second device from pre-stored modulation modes supported by the second device according to the identifier of the modulation mode supported by the second device carried in the first audio detection frame, and modulating the identifier of the modulation mode supported by the first device and the identifier of the modulation mode supported by the second device carried in the first audio detection frame using the modulation mode supported by the second device to generate the first audio detection feedback frame, sending the first audio detection feedback frame to the first device, and if the first device determines that the first audio detection feedback frame is correct, demodulating an audio data frame from the first device using a demodulation mode corresponding to the modulation mode supported by the first device.

9. The system according to claim 8, wherein the predetermined rule comprises combining a first modulation mode or a second modulation mode supported by the first device with a first modulation mode or a second modulation mode supported by the second device to obtain sets, setting a weight value for each set, and sequentially selecting the sets according to a magnitude of the weight values.

10. The system according to claim 9, wherein the first device is further configured for:

if the first device does not receive the first audio detection feedback frame, or if the first audio detection feedback frame is wrong,

modulating the identifier of the first modulation mode supported by the first device and an identifier of the second modulation mode supported by the second device using the first

a first determining module configured for determining whether a first audio detection feedback frame from the electronic signature token is received;

a second determining module configured for determining whether the first audio detection feedback frame is correct if the first determining module determines that the first audio detection feedback frame from the electronic signature token is received; and

a demodulating module configured for demodulating an audio data frame from the electronic signature token using a demodulation mode corresponding to the modulation mode supported by the electronic signature token, if the second determining module determines that the first audio detection feedback frame is correct.

16. The apparatus according to claim 15, wherein the predetermined rule comprises combining a first modulation mode or a second modulation mode supported by the apparatus with a first modulation mode or a second modulation mode supported by the electronic signature token to obtain sets, setting a weight value for each set, and sequentially selecting the sets according to a magnitude of the weight values.

17. The apparatus according to claim 16, wherein the modulating module is further configured for:

if the first determining module determines that the first audio detection feedback frame is not received, or if the second determining module determines that the first audio detection feedback frame is wrong,

modulating the identifier of the first modulation mode supported by the apparatus and an identifier of the second modulation mode supported by the electronic signature token using the first modulation mode supported by the apparatus according to the predetermined rule to generate a second audio detection frame, or

modulating an identifier of the second modulation mode supported by the apparatus and the identifier of the first modulation mode supported by the electronic signature token using the second modulation mode supported by the apparatus according to the predetermined rule to generate a third audio detection frame, or

modulating the identifier of the second modulation mode supported by the apparatus and the identifier of the second modulation mode supported by the electronic signature token using the second modulation mode supported by the apparatus according to the predetermined rule to generate a fourth audio detection frame,