CLAIMS

1. A method for separating in real-time voice from noise in an audio signal received via a receiver equipped with an audio sensor, said audio signal X having a length L, wherein the method comprises at least the following steps:
- separating the received audio stream into N frames X_N, wherein the method comprises iterating at least the following steps or each frame X_N :
- associating a tensor I_N containing information on the entirety of the audio stream with said frame X_N, the tensor I_N being determined from the tensor I_{N-1} associated with the frame X_{N-1} at the preceding iteration,
- transmitting each frame X_N to a module M_1, which is configured to generate a signal $X_N^{(1)}$, and transmitting said tensor I_{N-1} to a module M_3,
- transforming the signal $X_N^{(1)}$ by a module M_2 to generate a signal $X_N^{(2)}$,
- concatenating the signal $X_N^{(2)}$ and the tensor I_{N-1} by the module M_3 to generate a signal $X_N^{(3)}$ of 2F x T size, F being a given number of filters and T being a divisor of said length L,
- transmitting the signal $X_N^{(3)}$, to a module M_4 configured to generate a signal $X_N^{(4)}$ that is combined with the signal $X_N^{(1)}$,
- decoding the signal resulting from the combination by a decoder M_5, in order to generate a first signal $X_{N,0}$ being a voice signal and a second signal $X_{N,1}$ being a noise signal.

2. The method of claim 1, wherein the method comprises an encoding of the frame X_N by a 1D convolutional network, which provides the tensor $X_N^{(1)}$.

3. The method of claim 2, wherein the tensor I_N is determined from I_{N-1} according to :

$$I_N = I_{N-1} + \lambda(A_N - I_{N-1}),$$

λ being a gain factor and A_N being a tensor.

4. The method of claim 3, wherein the tensor I_N and the tensor A_N are of dimensions F x F.

5. The method of claim 3 or 4, wherein the tensor A_N is defined by:

$A_N = \left(\frac{X_N^{(2)} \cdot \left(X_N^{(2)}\right)^t}{\sqrt{T}} \right)$,

$X_N^{(2)} \cdot \left(X_N^{(2)}\right)^t$ being the matrix product of $X_N^{(2)}$ and of its transpose.

6. The method of any preceding claim, wherein $X_N^{(3)}$ corresponds to the concatenation of $X_N^{(2)}$ and C_N of dimension $F \times T$, C_N being defined by $C_N = B_N \cdot X_N^{(2)}$ and representing the matrix product of B_N and $X_N^{(2)}$, with $B_N = Softmax(I_{N-1})$.

7. A device for separating voice from noise in a received audio signal comprising a receiver equipped with an audio sensor, said audio signal X having a length L, wherein the receiver is configured to separate the received audio stream into N frames X_N, wherein the receiver is configured to iteratively process each frame X_N, the processing of the frame X_N comprising associating a tensor I_N containing information on the entirety of the audio stream with each frame X_N, said tensor I_N being determined from the tensor I_{N-1} associated with the frame X_{N-1} at the preceding iteration, wherein the receiver comprises at least the following elements to iteratively process each frame X_N:

- a module M_1 adapted to receive said frame X_N and to generate a signal $X_N^{(1)}$, the module M_1 being configured to transmit the tensor I_{N-1}, associated with the frame X_{N-1}, to a module M_3
- a module M_2 configured to transform the signal $X_N^{(1)}$ to generate a signal $X_N^{(2)}$ of 2F x T size, F being a given number of filters and T being a divisor of said length L, and to transmit the signal $X_N^{(3)}$ to a fourth module M_4,
- the module M_4 configured to generate a signal $X_N^{(4)}$ that is combined with the signal $X_N^{(1)}$,

the device further comprising a decoder M_5 configured to decode the signal resulting from the combination in order to generate a first signal $X_{N,0}$ being a voice signal and a second signal $X_{N,1}$ being a noise signal.

8. The device of claim 7, wherein the module M_{51} is configured to encode the frame X_N by a 1D convolutional network, which provides the tensor $X_N^{(1)}$.

9. The device of claim 8, wherein the tensor I_N is determined from I_{N-1} according to :

$I_N = I_{N-1} + \lambda(A_N - I_{N-1})$,

λ being a gain factor and A_N being a tensor.

10. The device of claim 9, wherein the tensor I_N and the tensor A_N are of dimensions F x F.

11. The device of claim 9 or 10, wherein the tensor A_N is defined by:

$A_N = \left(\frac{X_N^{(2)} \cdot \left(X_N^{(2)}\right)^t}{\sqrt{T}}\right)$,

$X_N^{(2)} \cdot \left(X_N^{(2)}\right)^t$ being the matrix product of $X_N^{(2)}$ and of its transpose.

12. The device of any preceding claim 7 to 10, wherein $X_N^{(3)}$ corresponds to the concatenation of $X_N^{(2)}$ and C_N of dimension $F \times T$, C_N being defined by $C_N = B_N \cdot X_N^{(2)}$ and representing the matrix product of B_N and $X_N^{(2)}$, with $B_N = Softmax(I_{N-1})$.