



Theme Name	Speech processing technology capable of extracting clear voice even in noisy environments with time variation
Organization Name	Prof. Tetsuya Shimamura, Graduate school of Science and Engineering, Saitama University
Technical field	IT

Overview

There are various noises in the everyday environment. Although noise removal technologies have advanced, it has been difficult to remove noise that has a large time change. In our laboratory, we have developed new noise reduction/removal technologies, such as in-frame processing method that works even with time-varying noise. This is effective not only for call quality such as telephone calls, but also for personal authentication and device control by voice, and is applicable not only to voice, but also to images. It can also support the use of deep learning.

We welcome companies that are willing to develop applications and businesses utilizing this technology.

Schematic Image

Speech processing technology capable of extracting clear voice even in noisy environments with time variation

Noise suppression technology using only current frame

【Features】

- A frame-based, real-time processing noise reduction method
- Noise suppression under various noisy environment including the conditions with time-varying noise
- Applicable to images as well as voice

【Applications】

【Example】 Applicable not only to call quality improvement but also to voice recognition



Recognition of instructions from voice



Telephone calls in noisy environment

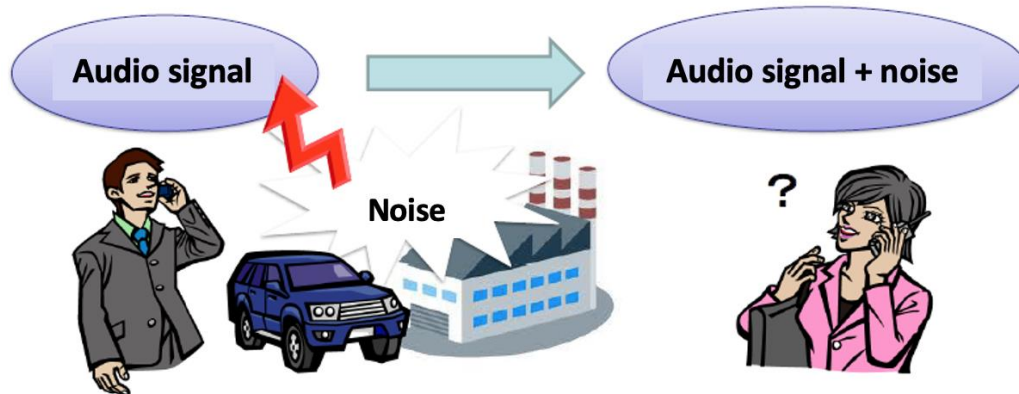


Security services using speaker authentication

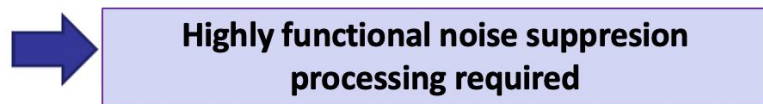


Background

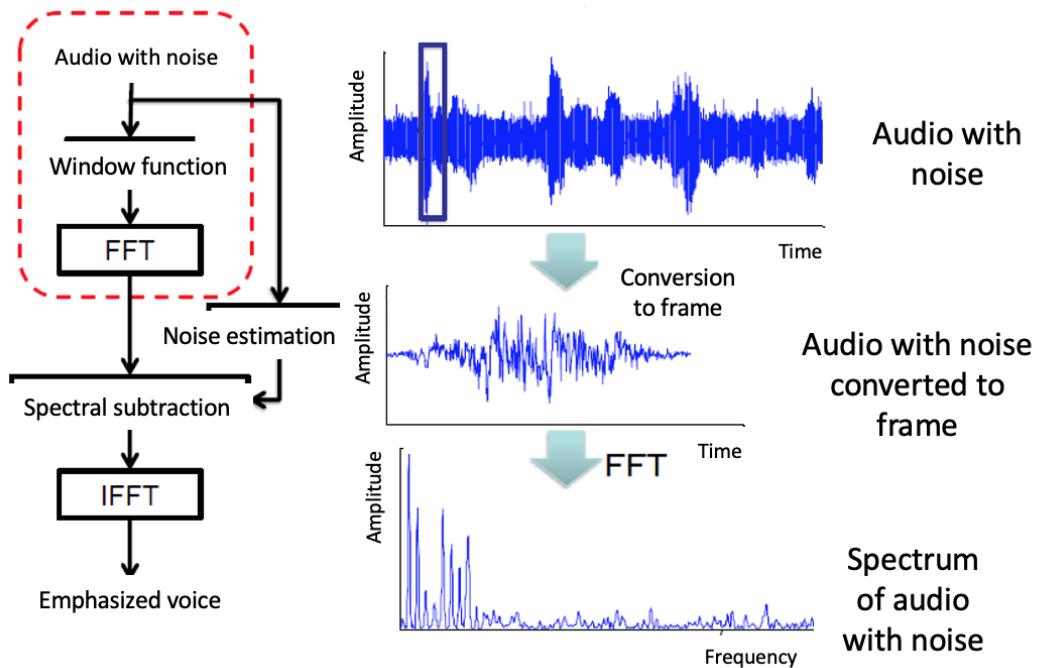
In everyday environment, various noises hinder smooth communication. For example, when the surrounding environment is noisy during a telephone call, noises are mixed into the audio signal, making it difficult to hear. In addition to this, clear audio is also required for voice-based personal authentication and device control.



Degeneration of audio quality at noisy environment



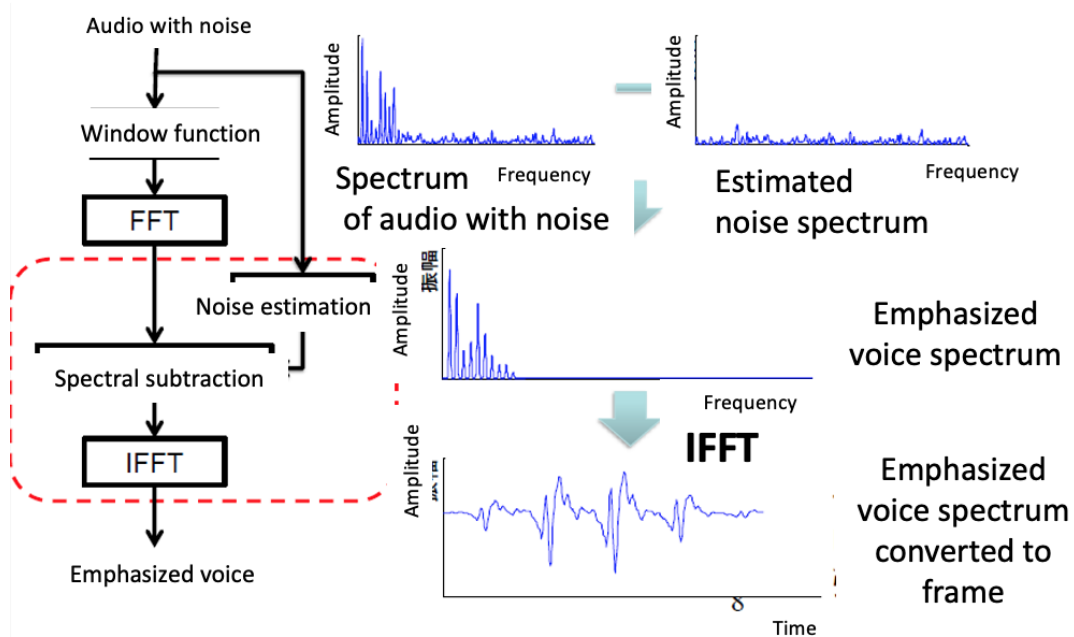
A well-known conventional noise reduction/removal technology is the spectrum subtraction method.



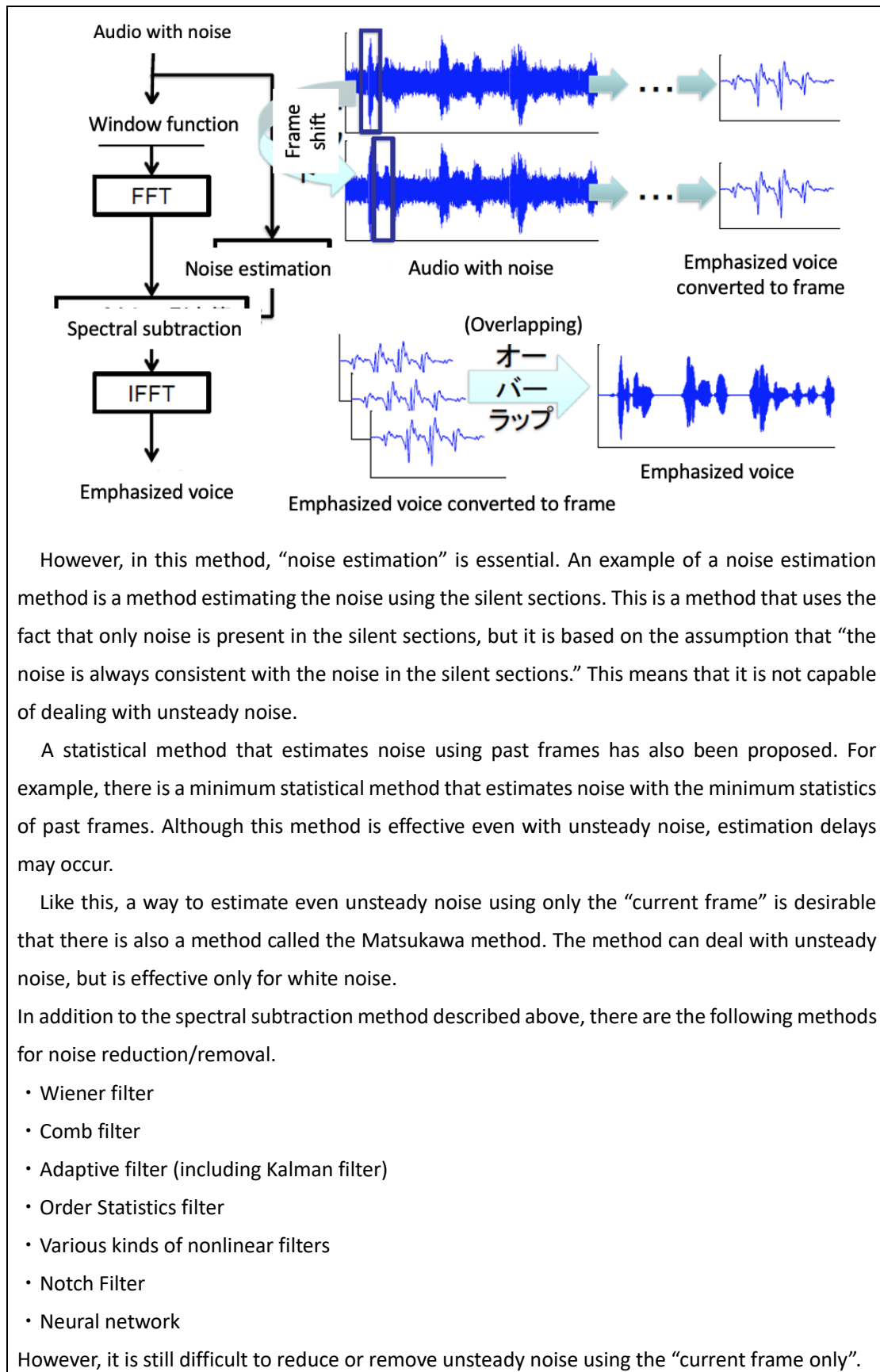


In this method, speech mixed with noise is first separated into frames with a time width of several tens of microseconds and then fast Fourier transform (FFT) is performed. In this state, the spectrum of noise and voice is mixed as shown in the bottom right of the figure above. Simultaneously, noise estimation is performed, which means the spectrum of the noises is estimated.

Then, the noise spectrum is removed from the mixed spectrum (spectral subtraction), and the remaining voice spectrum is inverted (IFFT) to extract the voice signal, as shown in the figure below.



This operation is repeated while shifting the frame, and the result can be overlapped and connected to obtain a clear voice with the noise reduced/removed.





Technical Content

The noise reduction/removal technology developed in this laboratory has the following characteristics.

- Voice can be effectively emphasized (signal-to-noise ratio improvement) in various noise environments.
- Noise reduction can be performed without tracking time-varying noise that is often found in the environment.
- Various frame-based noise reduction techniques can be performed in real-time processing.

For this, it is possible to extract a clear voice even in a noisy environment, and the technology can be applied not only to the audio quality of telephone calls, but also to voice control, recognition, and management under the real environment. Hence, it can be applied to personal authentication and device control by voice, which was difficult with the conventional noise reduction/removal techniques, and can be also applied to noise removal in an image as well as voice. In addition, it is applicable to the neural learning of both images and voice.

Strengths of the technology and know-how (novelty, superiority, utility)

The advantages of the noise reduction/removal technology developed in this laboratory are as follows.

- A noise suppression method that uses only the current frame.
- Can be used for various frame-based noise reduction techniques in real-time processing.
- Comprises of multiple methods and can effectively emphasize voice (signal-to-noise ratio improvement) in various noisy environments including the conditions with time-varying noise.
- Little distortion in musical noise (residual noise) and sound spectrum.
- Application to the neural network is possible.
- Applicable not only to voice but also to image.

Image of a Partner Company

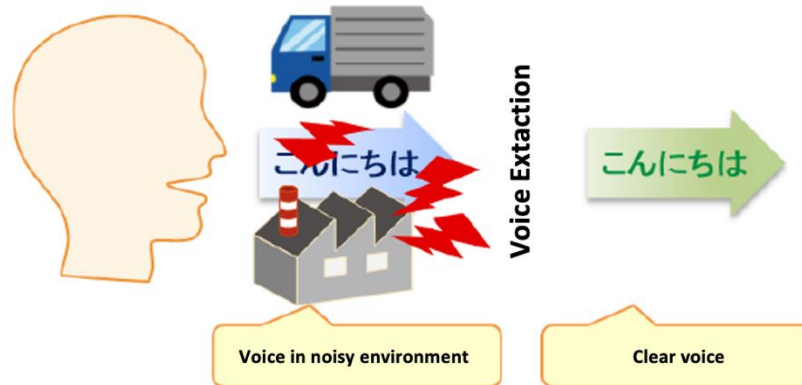
For example, we can cooperate with the following companies.

- 1) Companies related to mobile devices and their applications
- 2) Companies willing to develop voice-based control in automatic driving of cars, home appliances, and AI devices, and enter the voice application field.
- 3) Companies with voice security systems.
- 4) Companies that are interested in neural network applications.
- 5) Companies that want to remove noise under various circumstances (e.g., when working at near noise sources such as a factory or a construction site, or a drone.)
- 6) Companies that are interested in applying this technology to subjects other than voice, such as image sharpening.
- 7) Other companies that are willing to utilize and/or commercialize this technology.



Utilization of the technology and know-how (images)

In these ways, the noise reduction/removal technology developed in this laboratory is suitable for extracting a clear voice in a real-life environment with many noises.



Therefore, it is useful not only for improving the quality of voice calls in a real-life environment, but also for improving voice recognition performance.

- Voice recognition in mobile devices
- Instructions and conversations with electrical appliances and AI devices at home.
- Applicable to exporting audio to text in a real-life environment and uses that require precise voice recognition.
- Speech recognition and speaker recognition for automatic driving of a car
- Applicable to voice security systems.



Recognition of instructions from voice



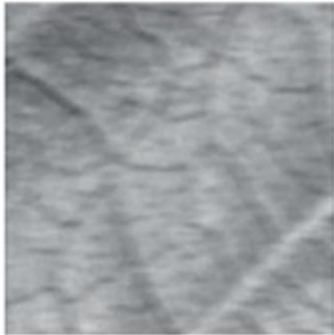
Telephone calls in noisy environment



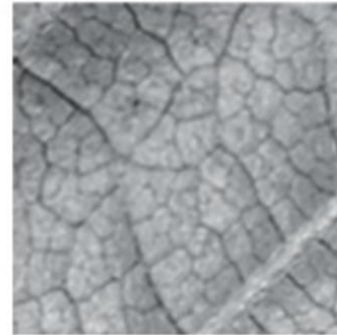
Security services using speaker authentication

Moreover, the noise reduction/removal technology developed in this laboratory is applicable to various fields where the signal and noise are separated (e.g., ocean, human body, living organism, music, etc.) and also to images.

Below is an example:



<Blurred image with noise>



<Recovered image>

Flow of the technology and know-how

If you are interested in utilizing or developing a product with this technology, please feel free to contact us. We will provide explanations with demonstrations.

Description of the technical terms

[Fast Fourier Transform (FFT)]

FFT is an algorithm that computes the discrete Fourier transform used in the frequency analysis of digital signals at a high speed on a computer. The inverse translation is called inverse fast Fourier transform (IFFT.)

[Comb Filter]

The comb filter is the filter whose frequency characteristics are the shape of spikes (the red line of the figure below). It is referred to as a comb filter because the filtered image has a “comb-like shape”. It may be used to extract voice (blue line in the diagram below), using the voice spectrum as an integer multiple of the fundamental frequency. The filter is also well-known for separating luminance and chrominance signals on color TVs.

