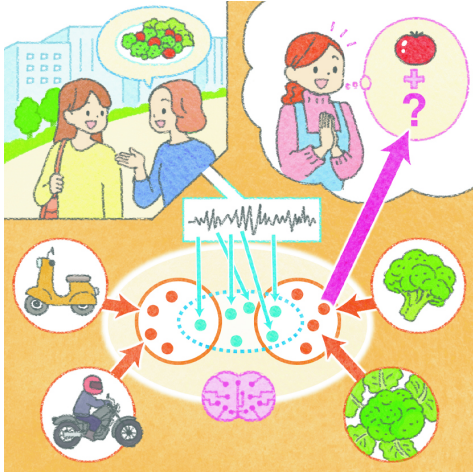


New signal processing to separate and extract speech by meaning



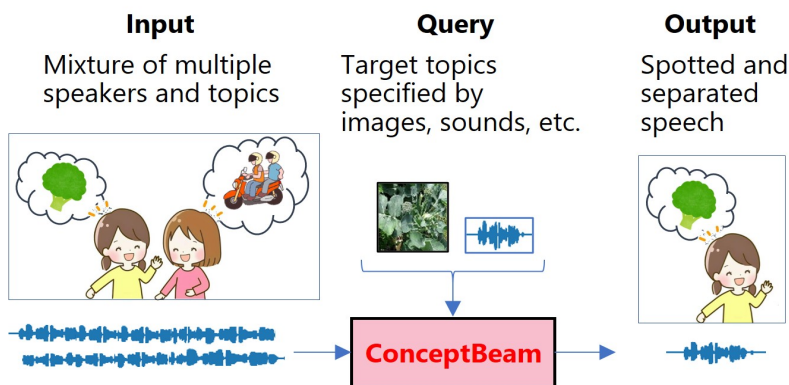
Background

The signal processing that separates and extracts the voice of a target speaker from an audio signal mixture is called sound source separation, and it plays an important role in speech interfaces. Traditionally, the separation has been based on physical cues such as sound direction, tone, and pitch.

Summary

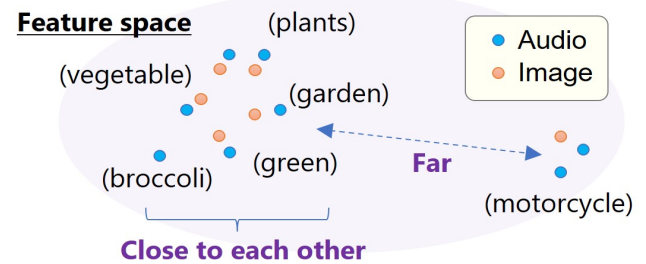
In contrast, this technology uses semantic content (concepts) as clues to identify the location of the voice speaking on the desired topic in a long acoustic signal, and furthermore, realizes a process to separate and extract only the desired speech.

ConceptBeam can detect and locate speech signals in which topics of interest are spoken and can extract only the speech of interest of the speaker of interest, even if the input signal is a mixture of voices from multiple speakers.



What's inside?

Our newly developed machine learning and signal processing techniques can measure semantic proximity between fragments of signals (input and query) in the feature space, regardless of the data type, and can extract only signal components close to the query from the input.



Features

- Targeted speech can be retrieved and separated without relying on speech recognition, which tends to be less accurate for signals with multiple simultaneous speakers and noises
- Specification of target information (concept specification) is possible by a speech, image, text, or their combination
- Supports concept specification across languages and information types (e.g., English speech can be retrieved by target specification with Japanese speech and/or images)

Future_benefits

Desired information will be more accurately and efficiently selected and utilized from the vast amount of media information on the Internet and records in the real world.

Exhibiting Company

NIPPON TELEGRAPH AND TELEPHONE CORPORATION

Contact

rdforum-exhibition@ml.ntt.com