ConceptBeam: Technology for separating signals by meaning

10

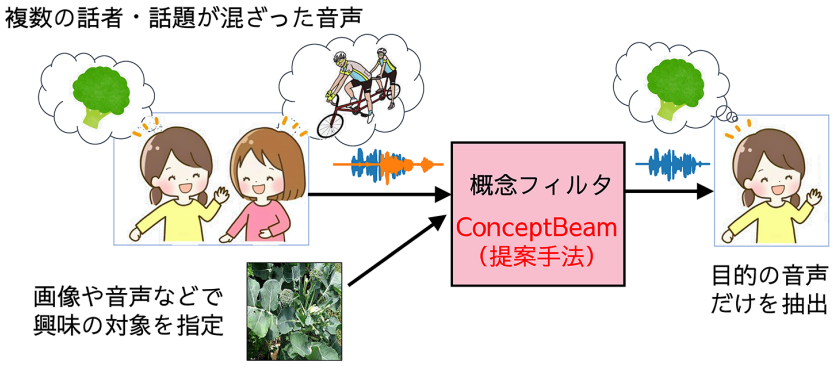
意味で音声を分離抽出する新しい信号処理技術ConceptBeam

興味のある話題に聞き耳を立てる

どんな研究	音や画像やテキストなどに表現され伝達される情報の「意味」を計算機上に表現し活用する技術を研究しています。本展示ではその一例として、複数の話者や話題が混在した音声信号から、音声、画像、テキストのどれかで指定した「意味」にマッチする話題の信号を取り出す技術を紹介します。
どこが凄い	複数話者の音声から目的の信号を取り出す方法として従来は専ら信号自体の性質(音の到来方向、信号源の独立性等）を用いる方法が研究されてきました。これに対しConceptBeamは話の内容(意味）に基づいて目的の信号を取り出せる、信号処理に意味理解を融合した世界初の技術です。
めざす未来	情報があふれる時代、有益な情報を抽出・選択することの重要性が高まっています。伝統的な信号処理やパターン処理に新たに意味処理を導入することで、多種の情報に対して興味のある情報を高速かつ的確に特定し、取り出し、活用できる社会の実現をめざします。

意味に基づいて音声を分離抽出する“概念フィルタ”

・複数の話者・話題が混ざった混合音声から 興味のある音声を分離抽出します。
興味の対象は画像や音声などで指定します。
・混合音声に対しては認識精度が低下しがちな音声認識を経由することなく、
意味の情報を直接用いて目的の音声を抽出できることが特徴です。



実験結果の例

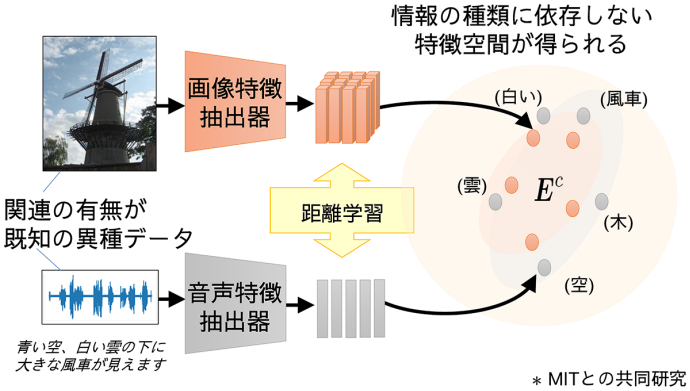
既存の技術を組み合わせる方法（手法１・手法２）に比べ 高い精度で音声を抽出できました。[1]

手法	概念の指定方法		精度評価値＊（数字が大きいほど良い）	
	画像	音声	2発話 2概念	4発話 2概念
手法１ 音声認識してから単語の情報で分離	✓		1.3 dB	-3.0 dB
		✓	1.1 dB	-1.4 dB
手法２ 音源分離してから信号を選択	✓		8.6 dB	1.0 dB
		✓	7.9 dB	-0.8 dB
ConceptBeam	✓		10.3 dB	4.0 dB
		✓	11.4 dB	3.8 dB

＊ 入力信号の重なり率が 50% のときのスペクトル歪みの改善度

仕組み① 概念の表現方法

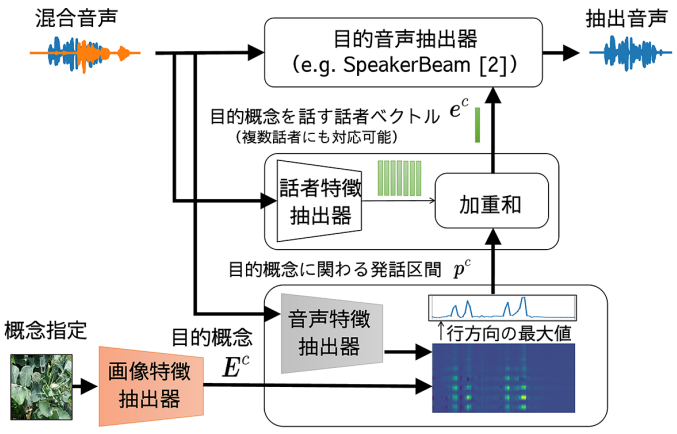
画像や音声などといった情報の種類に依存しない特徴空間を作り、その中のベクトルで概念を表現します。このような特徴空間は「関連の有無（例えば同時に生じているか否か）が既知のデータ」を使って構築できます。*



＊ MITとの共同研究

仕組み② 概念を用いたフィルタリングの方法

概念に関わる発話区間の話者ベクトルを推定し、音声抽出に利用します。



関連文献

[1] Y. Ohishi, M. Delcroix, T. Ochiai, S. Araki, D. Takeuchi, D. Niizumi, A. Kimura, N. Harada, K. Kashino, “ConceptBeam: Concept Driven Target Speech Extraction,” in Proc. *ACM Multimedia*, pp.4252-4260, 2022.
[2] M. Delcroix, K. Zmolikova, 木下慶介, 荒木章子, 小川厚徳, 中谷智広, “SpeakerBeam: 聞きたい人の声に耳を傾けるコンピュータ――深層学習に基づく音声の選択的聴取,” *NTT技術ジャーナル*, Vol. 30, No. 9, pp. 12-15, 2018.

連絡先

柏野 邦夫（Kunio Kashino）
メディア情報研究部 生体情報処理研究グループ