

# 21

## 会話の流れを読んでみんなの声をうまく認識

～頑健な音声処理を可能にする複数人発話区間推定技術～



### どんな研究

複数人の会話音声から、各話者が話した時間区間（発話区間）を精度よく推定する技術を紹介します。発話区間が分かると、例えば議事録に発言者を併記したり、対話ロボットにタイミングよく発言させたりできるほか、発話区間情報を活用した各話者の音声認識率の改善も可能です。

### どこが凄い

各話者の位置特徴量を精度よく表現する統計モデルに基づき高精度な発話区間推定が可能です。また、発話区間情報を使った安定的な各話者音声の強調技術や、発話区間や話者交替の情報を使った言語モデルにより、複数人会話における各話者の音声認識率を改善できます。

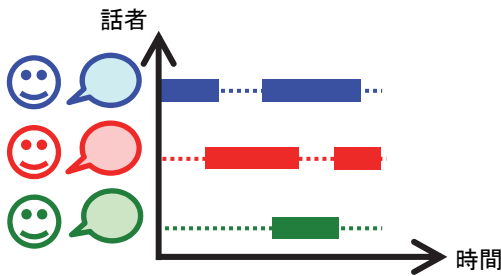
### めざす未来

現状の音声認識は、一人の人がスマホやロボットに意識して話しかける利用の仕方がほとんどです。これに対し「いつ誰が何を話したか」を高精度に自動推定する本技術で、人と人が話す自然な会話の中にコンピュータが入って来られるような音声入力インタフェースを実現します。

#### ◆発話区間推定とは：

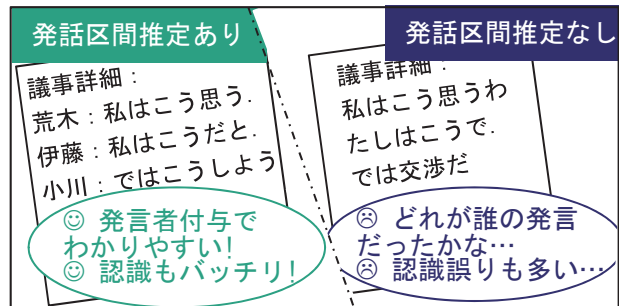
各話者が話した時間区間を推定する技術

（ダイアライゼーションと呼ぶことも）



#### ◆応用例：

##### ・会議録の自動作成

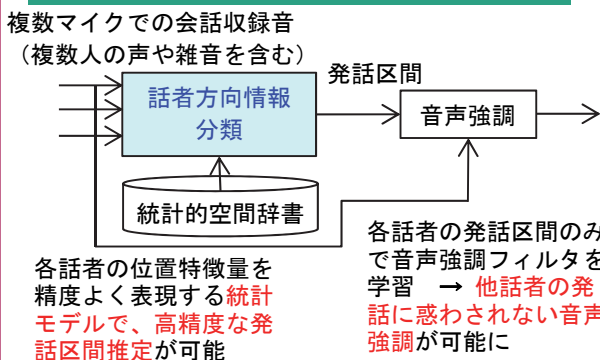


##### ・会話シーンでの難聴者支援

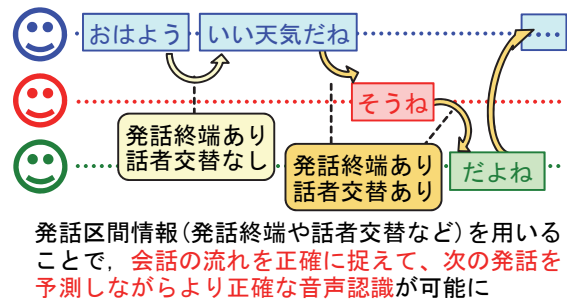
##### ・コミュニケーションロボットの制御

#### ◆発話区間推定技術と、その音声処理技術への応用：

##### 複数人会話の発話区間推定と音声強調



##### 発話区間情報を用いた会話認識用言語モデル



### 関連文献

- [1] N. Ito, A. Araki, M. Delcroix, T. Nakatani, "Probabilistic spatial dictionary based online adaptive beamforming for meeting recognition in noisy and reverberant environments," in *Proc. ICASSP2017*, 2017.
- [2] 芦川博人, 森岡幹, 小川厚徳, 岩田具治, 俣直弘, 小川哲司, 小林哲則, "複数人対話のための話者情報を用いたRNN言語モデル," *日本音響学会2016年秋季研究発表会講演論文集*, 3-Q-2, pp. 85-88, Sept. 2016.

### 連絡先

荒木 章子 (Shoko Araki) メディア情報研究部 信号処理研究グループ  
E-mail: araki.shoko(at)lab.ntt.co.jp