

NEWS RELEASE

平成 28 年 (2016 年) 9 月 29 日

複数話者の音声を同時認識する新しい音響信号処理技術を開発 音声認識による会話の文字起こしも可能に

大学共同利用機関法人 情報・システム研究機構 国立情報学研究所 (NII、所長：喜連川 優、東京都千代田区) 情報学プリンシプル研究系准教授、小野 順貴 (おの・のぶたか) の研究室と国立大学法人 筑波大学 (学長：永田 恭介、茨城県つくば市) 生命領域学際研究センター教授、牧野 昭二 (まきの・しょうじ) らの研究グループは、複数の機器で録音した複数話者の重なり合った音声から一人ひとりの声を分離し、これにより複数の音声を同時認識できるようにする新しい音響信号処理技術を開発しました。本成果は、米サンフランシスコで今月開かれた音声分野のトップ国際会議「INTERSPEECH」^(*1) で、小野研究室の NII 情報学プリンシプル研究系特任助教、越智 景子 (おち・けいこ) が発表しました。この技術により、会議のように複数の話者が同時に話すような状況でも、特別な機器を使わずに音声認識することが可能になります。

音声認識の性能は近年大幅に向上しつつありますが、会話や会議に代表される複数話者環境では、異なる話者の音声が重なるため、認識性能の低下が大きな問題となっていました。複数の話者の音声が混在した状況で個々の話者の音声を分離する「音源分離」という技術も研究されていますが、従来の技術では、複数のマイクロフォンで同期録音を行うためのマイクロフォンアレイと呼ばれる特別な機器が必要でした。これは、別々の録音機器で録音された信号は、録音開始時間が異なるだけでなく、サンプリング周波数 (音圧信号をデジタルの信号時系列に変換する際の周波数) も機器ごとに微小に異なっており、従来の音源分離手法はこうした信号を扱うことができなかったからです。

NII と筑波大学の研究グループは、別々の機器で録音され、同期がとれていない複数の録音信号を、録音後に同期させる新しい信号処理技術を開発しました。この技術と、小野研究室が開発した高速なブラインド音源分離^(*2) の技術を組み合わせ、複数話者の音声が混ざり合った会話を個々の音声に分離した後、音声認識を適用することで、複数話者環境の音声認識性能を大幅に向上することに成功しました。

本技術の実用例の一つとして、次のようなミーティングの自動議事録作成システムが考えられます。

- ① 録音：ミーティングの参加者は自分のスマートフォンを持参し、各自のスマホでミーティングでの会話を録音します。

NEWS RELEASE

- ② 同期：各参加者のスマホで録音された音声は同期されていないので、本共同研究で開発された信号処理技術で1ミリ秒よりも細かい精度で自動的に同期させます。
- ③ 分離：それぞれのスマホで録音された音声にはスマホの持ち主以外のものも混じっているので、ブラインド音源分離技術によって参加者ごとの音声に分離します。
- ④ 認識：分離された音声を、音声認識技術を使ってテキスト化します。

NIIでは、この技術をwebベースのGUIシステムとして実装する研究^{(*)3}も進めています。

これらの研究は、JSPS 科研費 JP16H01735^{(*)4}の助成を受けたものです。

以上

〈メディアの皆様からのお問い合わせ先〉

大学共同利用機関法人 情報・システム研究機構

国立大学法人 筑波大学

国立情報学研究所

総務部企画課 広報チーム

広報室

TEL:03-4212-2164 FAX:03-4212-2150

TEL: 029-853-2039 FAX : 029-853-2014

E-mail : media@nii.ac.jp

E-mail : kohositu@un.tsukuba.ac.jp

以上

本件はNIIと筑波大学が共同で発表するものです。NIIから文部科学記者会と科学記者会、筑波大学から筑波研究学園都市記者会を通じて各加盟メディアの皆様へ資料提供しているほか、NII、筑波大学それぞれから関係各メディアの方々に個別に本リリースをお送りしています。重複して配信される場合がありますことをご了承願いたします。

(*)1 「INTERSPEECH」： 音声科学と音声工学の国際学会「International Speech Communication Association (ISCA)」が主催する国際会議。本年は9月8日～12日に米サンフランシスコで開催。
<http://www.interspeech2016.org/>。

(*)2 「小野研究室が開発した高速なブラインド音源分離」： ブラインド音源分離とは、例えばスマートフォンや話者の位置情報などを用いず、録音された混合音声のみを用いてそれぞれの音源を推定する手法。小野研究室が開発した高速アルゴリズムの例は、以下の同研究室の資料「iPhoneアプリによるブラインド音源分離デモ」(http://www.onn.nii.ac.jp/demo_iphoneBSS.html) 参照。

(*)3 「webベースのGUIシステムとして実装する研究」： 詳細は、以下の小野研究室の資料「ウェブベースの議事録作成システム」(http://www.onn.nii.ac.jp/demo_multiASR.html) 参照。

(*)4 「JSPS 科研費 JP16H01735」： 基盤研究 (A) 「非同期分散チャンネルへ展開するアレイ信号処理理論の深化と実世界応用」。研究代表者：小野順貴。
<https://kaken.nii.ac.jp/ja/grant/KAKENHI-PROJECT-16H01735/>。