

# LINE、音声・音響信号処理分野で世界最大の国際学会「ICASSP」にて3本の論文採択

2022.02.17 技術情報

昨年に引き続き、NAVERとの共同研究7本を含む計10本の論文採択

音声認識や音源分離における性能の向上および実用性で評価

LINE株式会社（本社：東京都新宿区、代表取締役社長：出澤 剛）は、音声・音響信号処理分野におけるトップカンファレンス「ICASSP」にて3本の論文が採択されましたので、お知らせいたします。

ICASSP (International Conference on Acoustics, Speech, and Signal Processing)は、IEEE Signal Processing Societyが主催する音声・音響・信号処理分野における世界最大の国際会議で、2022年は47回目の開催となる非常に歴史が長く権威あるカンファレンスです。

LINEでは、昨年の「ICASSP 2021」においても7本の論文が採択されており<sup>\*1</sup>、昨年に引き続いてLINEで3本、NAVERとの共同研究7本を含め計10本の論文採択を達成いたしました。このたび採択された3本の論文は、5月に開催される「ICASSP 2022」にて発表を行います。

<sup>\*1</sup> LINE、音声・音響信号処理分野で世界最大の国際学会「ICASSP」にて国内トップクラスの7本の論文採択 <https://linecorp.com/ja/pr/news/ja/2021/3639>

## ■LINEが音声・音響信号処理分野の基礎研究として注力してきた音声合成技術、音源分離技術、環境音識別技術

LINEでは、AI事業を戦略事業の一つとして位置付け、AI技術の研究・開発およびAI技術を活用した事業の発展を加速させることを目的に、NAVERとの連携も行いながら、新たなAI関連サービス・新機能の創出を進めるとともに、それらを支える技術の基礎研究に注力しています。

基礎研究においては、機械学習を軸に、音声処理、言語処理、画像処理などに注力しています。音声・音響信号処理分野ではこれまで、高速かつ高品位な音声合成技術であるParallel Wave Gan、様々な音が混ざった音をそれぞれの音に分離する音源分離技術、音声からテキストを直接認識できるEnd-to-End音声認識技術、さらに、国際的なコンペティション「DCASE2020 Challenge(Task4)」にて世界1位を獲得した環境音識別技術などの研究を進めてまいりました。

## ■音声認識や音源分離のアルゴリズムのコンパクト化・高速化・学習効率化に関する研究が評価

今回の「ICASSP」では、音声認識、音源分離、自己教師あり学習の3つの関連領域における研究が採択されました。

まず音声認識関連では、非自己回帰型音声認識<sup>\*2</sup>の手法「Self-conditioned CTC<sup>\*3</sup>」のパラメータサイズの削減に関する研究が採択されました。ニューラルネットワークの各レイヤーに冗長性があることに着目し、同一の動きを持つ小規模ネットワークを再帰的に利用することで、パラメータを大幅に削減し小型化できることを示しました。「Self-conditioned CTC」と同等の性能を、従来の38%のパラメータ数で実現したことが評価されました。

次に音源分離関連では、アルゴリズム開発における出力の評価手法である「signal-to-distortion-ratio (SDR)」の研究で、従来の実装より10~100倍高速なSDRの算出を可能としたことが評価され採択されました。開発物については、GitHubでコードを公開しています<sup>\*4</sup>。

また自己教師あり学習関連では、音声データに関する汎用的な事前学習モデルに関する提案が採択されました。従来の自己教師あり学習では、単一の観点から定義した損失関数を用いていたのに対して、複数観点から損失関数を新たに定義することで、自己教師あり学習による学習データ整備の容易性に加え、より広範囲のタスクに適用可能な事前学習モデルを実現したことが評価されました。

<sup>\*2</sup> 非自己回帰型音声認識：ある文字を予測する際に、それ以前に認識した文字列を参照しない音声認識の手法。各文字を並列に予測することができるため、高速な認識処理が可能となる。

<sup>\*3</sup> Self-conditioned CTC：非自己回帰型音声認識の最新技術の一つ。ニューラルネットワークの中間層でも音声認識を行い、その結果を後段へフィードバックすることで精度の高い認識を行う。[Nozaki and Komatsu, Interspeech2021]

<sup>\*4</sup> [https://github.com/fakufaku/fast\\_bss\\_eval](https://github.com/fakufaku/fast_bss_eval)

## 採択された論文

### ●NON-AUTOREGRESSIVE ASR WITH SELF-CONDITIONED FOLDED ENCODERS

Tatsuya Komatsu

### ●SDR -- MEDIUM RARE WITH FAST COMPUTATIONS

Robin Scheibler

### ●SELF-SUPERVISED LEARNING METHOD USING MULTIPLE SAMPLING STRATEGIES FOR GENERAL-PURPOSE AUDIO REPRESENTATION

Ibuki Kuroyanagi and Tatsuya Komatsu

## ■今後もAI技術に関わる基礎研究を積極的に推進、既存サービスの価値向上を目指す

LINEが提供するAIテクノロジーブランド「LINE CLOVA」は、さまざまなAI技術やサービスを通して、生活やビジネスに潜む煩わしさを解消すること、社会機能や生活の質を向上させることで、より便利で豊かな世界をもたらすことを目指しています。現在、音声に関連する技術として「CLOVA Speech（音声認識）」「CLOVA Voice（音声合成）」、そしてそれらの技術を組み合わせたソリューションの提供も行っています。

たとえば「LINE AiCall」は、「CLOVA Speech」と「CLOVA Voice」、および会話制御の仕組みを組み合わせることで、ユーザーの要望に対してAIによる自然な対話応答を実現し、目的を達成するソリューションで、行政や飲食店舗などでの活用が進んでいます。また、昨年発表した「CLOVA Note」は、会議における会話を高い精度で認識し、議事録として記録・管理するためのアプリケーションです。長時間の録音データ解析にも対応した音声認識モデルを用いることで、高い認識精度を実現します。

今後も、AI技術に関連する基礎研究を積極的に推進することで、既存サービスの品質向上や、新たな機能・サービスの創出に努めてまいります。

LINEでは、今後も事業発展やサービスの価値向上のための取り組みを積極的に行い、コミュニケーションプラットフォームとして更なる成長・拡大の可能性を広げてまいります。

---

© LY Corporation