

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

2021.02.26 技術情報

NAVERとの共同研究2件を含め、LINEとNAVERで計14本の論文採択

音声合成の品質改善や、音声認識率の向上を達成した研究などで評価

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7本の論文が採択されたので、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、2021年は46回目の開催となる非常に歴史の長い権威あるカンファレンスです。今年は3600件超の投稿数があり、約1700件の論文が採択されています。採択された論文は6月にオンラインで開催される「ICASSP 2021」にて発表を行います。

■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。そして、「データ基盤開発」「データ分析」「機械学習」「AI技術開発」「基礎研究」を担う各チームが事業や担当領域を超えて連携することで、「研究＞開発＞事業化」のサイクルをスピードアップすることを目指しています。

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・音響信号処理分野においては、これまでParallel Wave GanというGPUを利用した高速かつ高品位な音声合成技術や、様々な音が混ざった音をそれぞれの音に分離し音声品質や音声認識率を高めることを目的にした音源分離技術、さらに、身の回りで起こる多種多様な音を機械で自動的に検出・認識する環境音識別技術の研究を進めてまいりました。

■音声合成の品質改善や、音声認識率の向上を達成した研究などが評価され採択

今回のICASSP2021では、音声合成関連として、敵対的生成ネットワーク（Generative Adversarial Network / GAN）^{*1}を用いた非自己回帰型音声生成モデル^{*2}であるParallel WaveGAN において、有声/無声の情報を利用して識別器を改良した研究が採択されました。単一の識別器を用いた従来のParallel WaveGANでは、識別器の表現力の制限、また学習の難しさから、特に複数話者コーパスを用いた際に品質が劣化する問題がありました。提案法では、有声音と無声音で音声の性質が異なることに着目し、それぞれの性質に合わせた識別器を設計することで、合成音声の品質を大きく改善しました。逐次的な研究ではあるものの、男女合計4名の話者を用いて大規模な主観評価実験を行い提案法の有効性を検証したことが評価され、採択されました。

また音源分離関連では、深層学習を用いない音源分離方式であるISS (Iterative Source Steering)法と深層学習を用いた音源モデルの推定法を組み合わせた新しい手法を提案し、従来のISSを超えて音声認識率が向上したことと、提案法が音源数に依存せずに適用可能なフレームワークであることが評価され、採択されました。

^{*1} 機械学習の生成モデルのひとつ。2つのニューラルネットワークを用いて学習を行い、入力されたデータから新しい擬似データを生成する。

^{*2} 過去に生成した音声に依存せずに、各時点の音声を生成するモデル。並列処理が可能であるため計算効率が高い。

このたび採択された論文

- ・ PARALLEL WAVEFORM SYNTHESIS BASED ON GENERATIVE ADVERSARIAL NETWORKS WITH VOICING-AWARE CONDITIONAL DISCRIMINATORS
R. Yamamoto, E. Song, M. Hwang, and J. Kim
- ・ TTS-BY-TTS: TTS-DRIVEN DATA AUGMENTATION FOR FAST AND HIGH-QUALITY SPEECH SYNTHESIS
M. Hwang, R. Yamamoto, E. Song, and J. Kim
- ・ END TO END LEARNING FOR CONVOLUTIVE MULTI-CHANNEL WIENER FILTERING
M. Togami
- ・ DISENTANGLED SPEAKER AND LANGUAGE REPRESENTATIONS USING MUTUAL INFORMATION MINIMIZATION AND DOMAIN ADAPTATION FOR CROSS-LINGUAL TTS
D. Xin, T. Komatsu, S. Takamichi, H. Saruwatari
- ・ SURROGATE SOURCE MODEL LEARNING FOR DETERMINED SOURCE SEPARATION
R. Scheibler, M. Togami
- ・ REFINEMENT OF DIRECTION OF ARRIVAL ESTIMATORS BY MAJORIZATION-MINIMIZATION OPTIMIZATION ON THE ARRAY MANIFOLD
R. Scheibler, M. Togami
- ・ JOINT DEREVERBERATION AND SEPARATION WITH ITERATIVE SOURCE STEERING
T. Nakashima, R. Scheibler, M. Togami, N. Ono

■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特化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インフラとして更なる成長・拡大の可能性を広げてまいります。
