会議レポート

音声音響信号処理の国際会議参加報告
― IEEE ICASSP 2015 ―

会議の概要

International Conference on Acoustics, Speech and Signal Processing（ICASSP, IEEE 信号処理ソサエティ主催）は，音声音響信号処理をはじめ，信号処理技術全般について幅広く議論する国際会議である．1976 年フィラデルフィアで第 1 回会議が開催されて以来，毎年，第一線で活躍する本技術分野の研究者が一堂に会し，最先端技術を議論する場として，重要な役割を果たしてきた．本技術分野の最新動向を一度に把握するには最良の会議といえる．

今年の ICASSP は，オーストラリアのブリスベンで，2015 年 4 月 19 日から 24 日に開催された（**図 -1**）．最新成果に関する 1,283 件の研究発表に加えて，チュートリアル, キーノートなどさまざまな技術講演が行われた．また今年から，新たな試みとして School of ICASSP という一連のセッションが ICASSP の参加者全員が聴講できる企画として実施され，各分野の重要な基本技術を分野外の研究者や入門者が学べる機会が設けられた．

ICASSP では，例年，オープニングセレモニーで，IEEE 信号処理ソサエティの活動報告に加えて，ソサエティの各賞の受賞者や新 IEEE フェローが発表される（**図 -2**）．今年（2015 年）日本からは, 2 件のベストペーパー賞（東工大，NTT）と 1 人の新フェロー（東工大）が紹介されていた．

最新の研究動向

ICASSP で議論されるテーマは非常に多岐にわたる．たとえば，音声言語処理，音響信号処理，音楽信号処理，信号情報処理，信号処理理論，機械学習，画像信号処理，センサアレイ信号処理，マルチメディア信号処理などが含まれる．以下，筆者が最も関心を持っている音声認識と音響信号処理の研究動向に焦点を当てて報告する．

図 -1　ブリスベンの会場周辺の様子

図 -2
オープニングセレモニーの様子

音声認識

スマートフォンによる音声認識サービスの普及や，ディープニューラルネットワーク（DNN）などの深層学習によるパターン識別技術の登場により，近年，音声認識技術は急速な発展を遂げている．ICASSP においても，深層学習を用いた技術の発表が多数を占めており，非常にアクティブな研究領域になっている（**図 -3**）．

音声認識の中で深層学習が最も利用されているのは，各短時間区間の音声がどの音素に対応するかの確率を推定する音響モデルの部分である．これに関して，今年も多数の報告があった．特に，従来標準的に使われてきた DNN に加えて，畳込みニューラルネットワーク（CNN），再帰型ニューラルネットワーク (RNN), Long Short Term Memory（LSTM）を併用して用いる研究が増え，着実な改善が得られていた．十分な学習データが利用できる場合の音声認識において，これらはすでに標準技術になりつつある．

一方，深層学習の枠組みに大量の学習データを与えればどんな環境でも高精度な音声認識が実現できるようになるかというと，少なくとも現時点ではそのように考えられていない．特に，収録環境の違いなどにより音響信号の特徴が大きく変化するような状況では，深層学習は

高い識別性能を発揮できなくなる．そこで，事前学習したニューラルネットワーク（NN）を用いながら，学習した環境とは異なる収録環境でも適切な音声認識を実現するために，少量の音声データを用いて追加学習を行う音響モデル適応の技術が重要な課題として多くの発表で取り上げられていた．また，近年，雑音や残響のある環境での音声認識（特に，マイクから離れた位置で人が話す状況を想定した遠隔発話音声認識）が盛んに研究されている．今年の ICASSP でも，深層学習に基づく音声認識を雑音除去や残響除去と統合的に用いる技術が多数報告されていた．これらの研究の評価には，CHiME Challenge や REVERB Challenge などの，近年実施されたチャレンジ企画の音声データが広く利用されていたことも今回の特徴の 1 つであった．

音響モデル以外では，文中での単語のつながりやすさを確率的にモデル化する言語モデルに RNN を用いる技術（RNN 言語モデル）の研究が多数報告されていた．従来の言語モデルと比べ，より長い文脈を考慮することができ，より高い単語予測精度が実現できるという．今後，音声認識の標準技術として広く利用される可能性が高いと思われる．

深層学習に関するもう 1 つの特筆事項として，マイクロソフトがオープンソース DNN ツールキット（Computational Network ToolKit : CNTK）をリリースし，チュートリアルで紹介していた．複雑な構造を持つ NN を簡単に構成できるなどの利点があり，今後，学術領域において広く利用されるようになる可能性がある．

音響信号処理

音響信号処理には，マイクロホンアレイ処理，音声強調，音源分離，音楽情報処理，音声音響符号化，音場再生など，多数のトピックが含まれている．全体を通じた動向の抽出は難しいが，あえて目立った動きを 1 つ挙げるならば，非負値行列因子分解（NMF）が基本ツールとして幅広く利用されるようになってきている．ベイズ的な生成モデルアプローチに基づく統計的信号処理の枠組みとよくマッチし，伝統的なマイクロホンアレイ処理とも相性がよく，さまざまな応用が提案されていた．

図 -3　混雑しているポスター会場

一方，NMF ほどではないが，深層学習を用いた信号分析技術も，着実に利用が広がっている．特に，雑音や残響でひずんだ音声からクリーンな音声を回復する音声強調では，Denoising Autoencoder（DAE）が，強力なツールとして実証されつつある．今後，音響処理の中でも，利用が拡大していくことはほぼ確実と思われる．

また，音響信号処理における特筆すべき活動として，2014 年に標準化された音声音響符号化方式 Enhanced Voice services（EVS）に関する特別セッションが，今回の ICASSP で企画されていた．EVS 規格は従来の携帯電話とほぼ同じ情報量で，従来の 4 倍の音声帯域を出力でき，大幅な音楽の品質改善も確認されている．将来的に，さまざまな音声・音響サービスで利用されるようになることが予想される．

来年の開催

来年の ICASSP は，2016 年 3 月 20 日から 25 日に，上海で開催が予定されている．本年同様，最先端技術に関する成果が，多数報告されることが期待される．アジアでの開催であり，日本からの参加が増えることも期待される．

（中谷智広／ NTT コミュニケーション科学基礎研究所）