

音声認識における頑健性 — 音響分析・音響モデル, なにが課題か —

コーディネータ: 北岡教英 (名古屋大)

話題提供者: 藤本雅清 (NTT 研究所), 滝口哲也 (神戸大),
大淵康成 (日立), 中村 哲 (NiCT/ATR)

概要 音声認識実用化に向けて「頑健性」の必要性が叫ばれて久しい。頑健性を指向し、主に音響分析・音響モデルの側面から、いまなにが行われ、なにが課題とされているのかを考えてみたい。

Robustness in Speech Recognition — What is needed ? —

Cordinator: Norihide Kitaoka (Nagoya Univ.)

Panelist: Masakiyo Fujimoto (NTT Labs.), Tetsuya Takiguchi (Kobe Univ.),
Yasunari Obuchi (Hitachi Ltd.), Satoshi Nakamura (NiCT/ATR)

Abstract Robustness has been thought as very important for a long time to apply speech recognition systems to the real world. Here let us consider what is the 'next step' to improve the robustness from the viewpoints of acoustic analysis and acoustic models.

音声認識技術の実用化のためには、様々な意味での「頑健性」が求められることは共通の認識であろうと思う。頑健性には多様なレベルが考えられ、マイクロフォン(音のセンシング)から応用システムの工夫まで、考慮すべき点は多い。特に後者に近づくほどに、「音声認識が100%でなくとも(あるいは非常に性能が悪くとも)使って行けるアプリケーション」を開発するために頑健な処理を考案するアプローチと言え、実際に最近の国際・国内の学会発表においてこれらの意味で頑健性という言葉が用いられることが増えている。

こういったアプローチは確かに有効である。しかし前段の音響の段階での頑健性が揺らぐ中で、安定した頑健性を後段で維持するのは難しいことは容易に想像できる。車内からPC環境、ロボットなどアプリケーションが多様化し、言い替えば雑音・残響・ハンズフリーといった多様化に頑健さを持たぬ音響段階に後段で対処するというのは明らかに困難である。

実用化に向けた課題や取り組みについては、本研究会の10月に行われるデベロッパーズフォーラムにおいて多くの参加者が集まり、幅広い話題について活発な議論がなされていることから多くの興味を持たれていることがわかる。

そこで今回は、より焦点を絞り、またいままでは使えなくとも将来性のある技術も見据えた議論が

できることを目的として、音響分析・モデルについて頑健性を考えてみたい。

話題を提供いただく方として4名をお願いした。NTT研究所の藤本氏は、以前から雑音抑圧や雑音下での音声区間検出(VAD)で活躍されている。特に現在は非定常雑音下でのVADおよび雑音抑圧に基づく音声認識に注力されている。神戸大学の滝口氏は、かつてから加法性・乗法性雑音下での音声認識のための音響モデルの研究をされ、最近では斬新な特徴抽出法の研究も精力的に行われている。日立中央研究所の大淵氏は、最近では音声特徴量の研究を行いながらも常に実用化を視野に入れた地に足を付けた研究開発が行われている。そして中村氏は、ATRなどにおいて長く研究を続け、携帯電話を通じた音声翻訳システムを実用化に導いた。この間、氏が中心的に研究を行ってきたのが音響分野であると認識している。やや我田引水的ではあるが、コーディネータが主査を務める雑音下音声認識評価WGの立ち上げ者かつ初代主査であり、以来本WGでは活動期間中、常にフロントエンドの頑健性を大切に考えて来た。

これらの方々に、ご自身の経験・主義主張(?)を中心に、現在の動向なども交えながら話題提供をいただき、場におられるみなさまから活発に御意見をいただいで今後の研究活動に刺激を与えられれば幸いである。