

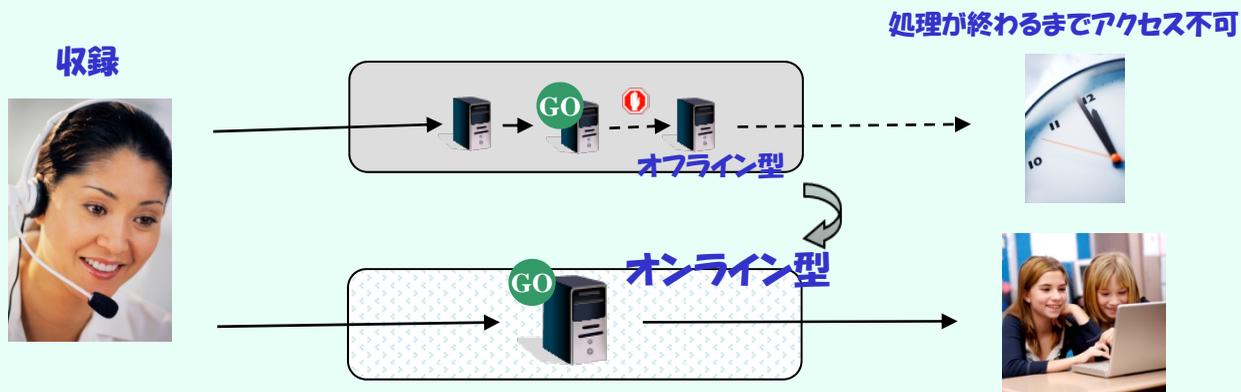
会話の記録をさくさく作る —オンライン音声言語処理技術—

背景・課題：音声認識などの多くの音声言語処理技術には、処理結果を入力と並行して逐次的に出力し、利用可能とするオンライン処理が求められます。しかし従来、オンライン処理においては、長い時間をかけた解探索や、モデル適応/再認識アプローチ等が適用できないため、認識精度の確保は困難でした。

アプローチ：重み付き有限状態トランスデューサ(WFST)に基づく音声認識技術や、時間発展を考慮したモデリング手法、高度なモデル学習法、逐次同期処理などにより、高精度なオンライン音声言語解析を実現します。

到達点：「会話の中で一人一人の声を取り出します～オンライン発話抽出技術～」で紹介される技術と結合することで、複数人の会話をモニタリングし、音源分離・音声強調・発話検出・音声認識・メタ言語情報提示の一連の処理を全てオンラインで実現できるようになりました。

概要：オンラインでも高精度な音声言語処理を提供する要素技術・理論的背景を紹介



オンライン音声言語処理を支える要素技術

高精度音響モデル学習法dMMI^{※1)}

高精度言語モデル学習法WGCLM^{※2)}

高度なモデル学習技術により、オフライン型に匹敵する高精度なオンライン処理を提供します。

音声検出技術DIVIDE^{※3)}

追従型モニタリング

解析対象の時間変化をモデリングすることは、オンラインでの動作に加え、精度の向上にも繋がります。

※4) WFSTに基づく高速な

汎用音声言語処理系SOLON-GT^{※5)}

複数のモデルを柔軟に統合することで、高速な処理を実現します。

処理結果に即座にアクセス

会話のモニタリング
テレビ会議支援システム
人の会話に参加するコンピュータ等

の実現をサポート！

※1) differentiated Maximum Mutual Information ※2) Weighted Global Conditional Log-linear Model ※3) Dynamic Integration of Voice Identification and DE-noising
※4) Weighted Finite-State Transducer ※5) Spoken-language-processing Oriented Library based on Optimized transductive Networks with Generalized Transductions



関連文献

- T. Oba, et al., “A Comparative Study on Methods of Weighted Language Model Training for Reranking LVCSR N-best Hypotheses,” Proc. ICASSP, 5126–5129, 2010.
- M. Erik, et al., “Discriminative Training based on an Integrated View of MPE and MMI Margin and Error Space,” Proc. ICASSP, 4894–4897, 2010.
- S. Watanabe, et al., “Predictor–Corrector Adaptation by using Time Evolution System with Macroscopic Time Scale,” IEEE trans., ASLP. vol. 18, issue 2, 395–406, 2010.

連絡先

大庭 隆伸 (Takanobu Oba)

メディア情報研究部 信号処理研究グループ

oba@cslab.kecl.ntt.co.jp