

衆議院の新会議録作成システムにおける京都大学の音声認識技術の導入

2011年05月12日

衆議院の新しい会議録作成システムにおいて、京都大学・学術情報メディアセンターの河原達也教授らの研究開発による自動音声認識技術が導入された。約一年間の試行を経て、本格的に運用されることになった。

明治23年(1890年)に我が国に議院が設立されて以来百年以上にわたり、会議録の作成は手書き速記によって行われてきた。今世紀になって衆参両院において速記者の新規採用・養成が停止され、新たな会議録作成方法が模索された。様々な検討をふまえて、衆議院において音声認識技術を用いたシステムが採用された。世界的にみても、国会の審議音声の直接認識するシステムは初めての事例である。

このシステムでは、原則すべての本会議・委員会の審議において、発言者のマイクから収録される音声を自動音声認識により書き起こし、会議録の草稿を生成する(図1参照)。なお、音声認識には一定の誤りが不可避な上、話し言葉の発言を忠実に書き起こしても、会議録にならないので、速記者・校閲者の役割がなくなるわけではない。

音声認識技術はこの十年余りの間に進歩を遂げて、最近では携帯電話を用いた情報検索や自動翻訳などのサービスなどに導入されている。ただし、国会の委員会審議のような自発性の高い、人間どうしの自然な話し言葉音声を高い精度で認識できるものはなかった。

河原教授らは、まず衆議院の審議音声と忠実な書き起こし(実際の発言内容)からなるデータベース(=コーパス)を構築し、会議録の文章との違いを統計的に分析し、モデル化を行った。その結果、「えー」や「ですね」などの冗長語の削除を中心に、約13%の単語で違いが見られた(図2参照)。次に、この統計モデルに基づいて、大量の会議録テキスト(過去10年以上の分;約2億単語)から、実際の発言内容を予測するモデル(=言語モデル)を構築した。さらに、これと音声を照合することで、大量の審議音声(約500時間分)から音声パターンのモデル(=音響モデル)を構築した(図3参照)。これらは半自動的に追加学習・更新が可能であるので、今後総選挙や内閣改造があっても話者集合の変化を反映し、持続的に性能を改善していくことができる。

この音声認識技術は、一般競争入札を経て新会議録作成システムの開発を担当したNTT(NTT東日本、NTT研究所など)のシステムに組み込まれる形で導入された(本学産官学連携本部を通じてライセンス供与)。

昨年度の試行において性能評価を行ったところ、会議録と照合した音声認識結果の文字正解率は89%に達していた(デモあり)。これを、速記者が専用エディタで修正・編集することにより会議録原稿を作成するシステムの有用性が検証され、本格的な運用となった。

会議録作成以外の今後の展開としては、講演や講義などを対象とした字幕付与への取り組みを行っていく予定である。

この研究開発は、科学技術振興機構(JST)の戦略的創造研究推進事業(CREST)や、総務省の戦略的情報通信研究開発推進制度(SCOPE)などの国の競争的研究資金制度の支援を受けて行われた。

河原教授は、5月13日に大阪で開催される関西サイエンスフォーラムで、このシステムを含む音声認識技術の展望について講演する予定である。また、7月にパリで開催される国際速記連盟(Intersteno)の会議でも本研究開発について講演するよう依頼されている。



図1 システムの概要

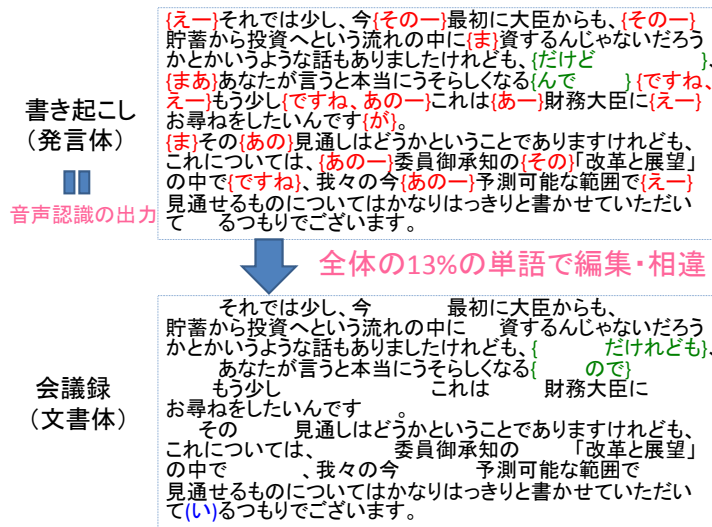


図2 書き起こし (発言体) と会議録 (文書体) の相違の例

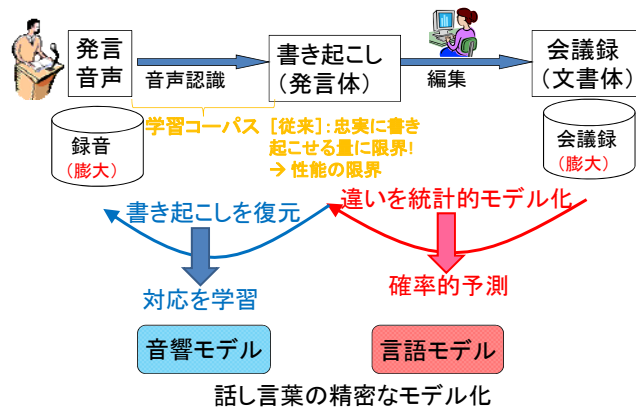


図3 本システムにおける音声認識技術の概要

* 科学技術的観点からの取材は個別に受付可能
河原達也教授 (E-mail: kawahara@i.kyoto-u.ac.jp) まで