

AI を活用した音声認識技術による自動文字起こし及び自動要約の実証実験

令和元年 7 月 31 日
総務省行政評価局
オフィス改革・働き方改革検討チーム

1 実証実験の概要

(1) 目的

音声認識・自動文字起こし技術及び自動要約技術について、中央省庁における会議や日々の打合せにおいて作成される議事録や議事要旨における作業の効率化を図る観点、また、その他どのような業務に活用することで業務効率化が可能なのか、その実用化の可能性について、これらの技術を試行的に活用することにより、実証的検証を実施するもの。

(2) 期間

平成 31 年 3 月 4 日から令和元年 6 月 30 日まで

(3) 使用ツール

① Smart 書記

マイクから集音した音声を、人工知能（AI）を活用した音声認識技術により自動でテキスト化する。

② QuickSummary

ユーザーが文字量（要約率）を設定すると、その分量に応じて対象となる文章(テキスト文書)を AI が要約を行う。

(4) 内容

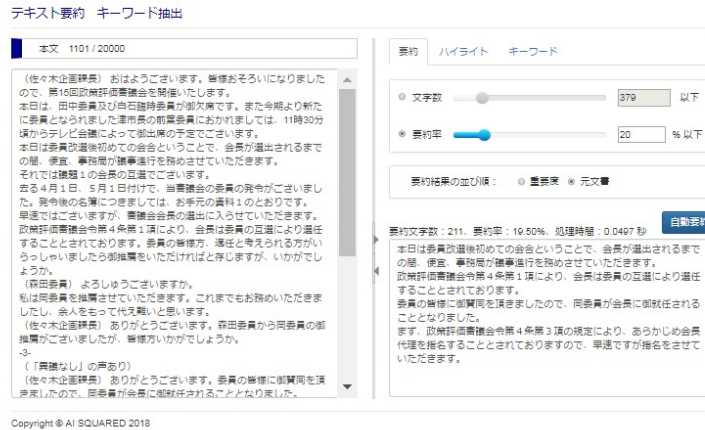
① 音声認識技術による自動文字起こし

行政評価局が主催する会議の議事録・議事要旨作成のほか、日常業務における打合せの議事メモの作成業務への実用可能性等を検証するため、これらの会議・打合せの音声データを利用して「Smart 書記」により自動文字起こし（テキスト化）を行い、その精度等を検証する。



② 自動要約

行政評価局が主催する会議の音声データ及び議事録（テキスト文書）を用いて、「QuickSummary」により要約を行い、議事要旨としての利用を想定した要約率ほどの程度（〇％）とするのが妥当なのかを検証する。



2 実証実験の結果

(1) 音声認識技術による自動文字起こし

行政評価局では、業者の協力を得て、自動文字起こしに必要な集音機器等を活用し、同局が主催する会議（政策評価審議会、同部会）ほか、日常業務における打合せの音声を集音し、Smart 書記を使用して自動テキスト化を実施した。

その結果については、下表のとおりで、マイクとの距離、スピーカーの声の性質、ボリューム、スピード、話す顔の向きなどの集音環境により、その精度等は大きく影響を受けることが分かった。

単一指向性の精度が高いマイクを利用して原稿を読み上げるように喋る場合には、全体的にある程度理解できる内容となっている。

一方、複数人の会議を一つのマイクで集音した実験では、マイクとの距離や声のボリュームによっては、そもそも声が拾えないなどの結果もみられた上、声を拾っても、正確なテキスト化はできなかった。

また、いずれの場合にも、前もってシナリオが用意されていない打合せなどは、個人の話し癖や感嘆符などが入りやすいなど、文字起こしが正確になされていないケースが見られた。

既に知事記者会見で会見要旨の作成に実用化されているように、記者会見など原稿を元に、正確に伝えることを意識して話される言葉を、相応の集音環境で録音する場合には、ほぼ実用に耐えられるレベルのテキスト化が、技術的には実現しているということが、今回の実証実験でも認められた。

一方で、マイクの集音性能によるところも大きいですが、会議などでの話し言葉の場合には、認識率は相当程度落ちることが確認された。

今回、実証実験を行った会議（審議会）や部内打合せにおいては、集音において十分な環境を実現することができなかったため、残念ながら、実用化に足りるテキスト化は行え

なかった。

注) Smart 書記や QuickSummary は、本来、クラウド上のサービスとして提供されるものであるが、総務省の LAN 環境においては、現時点では、本クラウドサービスの利用が困難であるため、今回、実験用のスタンドアロンの端末を利用して実験を行った。音声認識を行うためには、スタンドアロン端末を話者の人数分用意することが必要であったが、複数端末を確保することができなかつたため、限定的な実験となった。

《業者からの補足情報》

利用する音声認識エンジンによっては、話される内容に応じたチューニングも可能で、「言語モデル」「音響モデル」の最適化を行うことにより精度は確実に向上する。

デバイス	使用した場面	集音環境	評価
 無線マイク	打合せ	会議室	<ul style="list-style-type: none">• ところどころ正しく文字起こしできていない箇所はあるが、ある程度理解できる内容となっている。
 ICレコーダー	行政評価局が主催する会議	会議室	<ul style="list-style-type: none">• 一度 ICレコーダーに録音したものを自動文字起こしシステムにかけたが、周辺環境からのノイズ混入が録音により顕著化するために、認識が困難なケースがある。
 卓上マイク	打合せ	会議室	<ul style="list-style-type: none">• 無線マイクと ICレコーダーと比較すると、正確に文字起こしできている。

(2) 自動要約

行政評価局では、従来の方法で当局が作成した議事要旨と、「QuickSummary」を利用し、当局が速記録をもとに作成した議事録をもとに要約したものとを比較し、その適切な要約率や内容等を検証した。

その結果、あらかじめ議事録が整理されたものであるという前提の上では、要約率を30～40%程度に設定すると、ある程度必要な情報が盛り込まれた理解できる要約となっていることが分かった。

また、Smart 書記や QuickSummary を活用すれば、音声認識から一気に自動要約を行うこと自体は、技術的に可能であるが、今回の実証実験においては、集音の段階で適切な録音環境が実現できなかったこともあり、音声からダイレクトの要約の作成は行えなかった。

3 まとめ

今回の実証実験を通して、AI を活用した音声認識技術による自動文字起こしと自動要約の性能については、必要なシステム環境、集音環境が実現可能であるという前提に立てば、一定程度、実用に耐えられる水準にあるのではないかと考えることが可能と思われる。

特に、音声認識技術による自動文字起こしについては、適切な集音環境等が整えば、十分に実用可能なレベルにあると考えられ、役所の業務の効率化に大いに資するのではないかとと思われる。

その一方で、あらかじめ用意された発言要旨を元に相当程度明瞭に発言される場合には音声認識率が高いものの、話し言葉で自由に発せられる音声については必ずしも認識率が低いなど、使う場面によって向き・不向きがあることも分かった。実用段階においては、この点を加味して利用する必要がある。また、特に集音については、話者の数に応じた単一指向性のマイクを用意するなど、一定の前提が必要となることもわかった。

《業者からの補足情報》

複数人で行われる会議音声のテキスト化については、集音時点での対処や、音声認識エンジンによるテキスト化段階での対処が考えられる。集音機器による対処とは、特定のマイクからの複数人の発話音声を話者ごとに分離し、独立して集音する技術の活用である。また、音声認識エンジンによっては、単一の音声ファイルを入力とし、音声分離技術で複数話者に分割することも可能である。

自動要約については、今回の実証実験では、適切な集音を実現することができなかったため、音声データからダイレクトの要約は困難であったが、速記録を元にした議事録(正確なテキストデータ)が用意されれば、相当程度の精度の高い要約が可能であることが確認できた。議事録が作成される予定の会議で、速報のために作成される議事要旨であれば、自動要約技術による議事要旨の作成も十分可能ではないかと考えられる。(ただし、今後、集音技術も含めた音声認識技術がさらに進化すれば、議事録自体の速報性が高まると考えられることから、議

事要旨について投資を行ってまでシステムを活用する必要性については、慎重な検討が必要である。)

《業者からの補足情報》

今回の実証実験では、速記録を元にした議事録そのままを抜粋する形での要約の検証に留まった。ただ、理想となる要約の教師データを一定量準備することを前提に、理想要約の結果に近い文書を作成するという形での要約も可能となってきている。

セキュリティ上、総務省のネットワークからクラウドサービスである自動文字起こし機能と自動要約機能をダイレクトに使用することが困難であるといった実際の業務に活用するためには解決すべき課題も見つかった。今後は、これらの機能の実用可能性を検討の中で、セキュリティポリシーとの関係整理についても、引き続き検討が必要と考えられる。