# 会議録テキストを用いた議会映像配信への字幕付与

河原達也（京都大学）

ブロードバンド通信やスマートフォン等の普及によって、マスメディアやソーシャルメディアにおいて動画の配信・視聴が一般的になっている。国会や地方議会の公式の記録は、速記者等によって作成された会議録であるが、現在では多くの議会の審議映像がライブ配信され、後でアーカイブにオンデマンドでアクセスできるようになっている。しかし、審議映像に字幕を付与することはまだ難しく、多くの場合提供されていない。自動音声認識技術は実用的なレベルになりつつあるが、一定の誤りが不可避である。この報告では、議会の映像配信に公式の会議録を使用することが可能か、またどのようにしてそれを実現できるかについて述べる。

**議会映像配信と字幕付与の現状**

Voutilainen ら[1]が EU 21 カ国で実施した調査によると、本会議の映像はこれらのすべての国で公開されており、委員会審議も大半の国で提供されている。公式の会議録はすべての場合に提供されているが、字幕が提供されているのは約半数の国だけである。なお本報告では、アクセシビリティ向上のために、一定の長さの発話や文の単位で音声と時間同期しているものを字幕と呼ぶ。音声認識を含むいくつかの方式が採用され、記録部と別の部門で作成される場合もある。

日本の国会では、衆議院・参議院ともに、委員会を含むすべての会議が映像配信されているが、2024 年 8 月に参議院でライブの自動字幕付与が開始されるまでは、付随するテキストは一切提供されていなかった。このライブ字幕は音声認識によって生成されており、一定の誤りが含まれている。日本には都道府県・市区町村など約 1800 の自治体がある。そのうち、約 1000 の自治体が議会の映像配信を行っている。そのうちの約 3 分の 1 は、コストの問題から YouTube チャンネルを利用している。しかし映像配信に字幕を付与している自治体はごく僅かである。

人手による字幕付与は非常にコストがかかるが、音声認識による自動字幕は一定程度の誤りが不可避である。したがって、自動字幕はライブの字幕付与に用いることはできるが、永続的なアーカイブには適切でない。その反面、公式の会議録は多くの場合提供されているが、音声と時間的に対応していない。

## 公式の会議録を審議映像の字幕に使用可能か？

公式の会議録は、会議における発話と逐一合致しているわけでないので、字幕には不適切であると考えられてきた。速記者は、発言の内容を変えることなく、可読性の向上のために、会議録を作成・編集してきた。これは、口頭で話された内容をテキストに変換する過程で必要なものであり、フィラーや冗長な単語の削除、文法的誤りや口語的表現の修正、一部のフレーズの並び替えなどが含まれる。日本の国会と欧州議会における編集過程の詳細な分析は、河原[2]によって行われている。

それによると、過去数十年間で、編集の比率が大幅に（平均 16.3%から 9.9%へ）減少している。これは、口語的な表現が受け入れられるようになり、会議録が以前よりも逐語的になったことを意味する[3]。この傾向は、映像配信の普及やソーシャルメディアにおける参照に影響されている可能性がある。このことは、映像配信に公式の会議録を利用できることを示唆している。 実際、テレビ番組の字幕においても、フィラーの削除や文法的誤りの修正など、最小限の編集が行われている。

## 公式の会議録を映像配信の字幕に用いる方法

公式の会議録ができるまでには時間がかかるため、ライブの字幕に用いることはできないが、アーカイブ（オンデマンド）の映像配信の字幕に使用するのは合理的と考えられる。しかし、実際に字幕に供するためには、何時間にも及ぶ長い会議の音声・映像を、長い会議録テキストと発話や文レベルで時間同期する必要があり、その作業は面倒である。ある事例では、1 時間の音声について手作業で行うとほぼ 1 日要している。この問題に対して、音声認識技術を用いることにより、以下の手順のように簡便に解決することができる。

(1) 会議音声に対して音声認識を行い、発話または文の時間情報（タイムスタンプ）を含む書き起こしを作成する。時間情報の取得方法は音声認識システムに依存する。
(2) 音声認識で作成した書き起こしと公式の会議録テキストとの文字単位の対応付け（アライメント）を行う。
(3) 音声認識で作成された書き起こしの時間情報を会議録テキストの対応する部分にコピーする。
(4) 会議録テキストを、長さと言語的境界の観点から字幕に適切な単位に分割する。

ステップ(2)の対応付けは、音声認識の誤りが 20%程度でも可能であり、字幕と音声の時間的なずれも一定程度は許容できることが実験的に確認されている。ただし、不

規則な事象等により、音声のかなりの部分が会議録に書き起こされていない場合は対応できない。

### 地方議会のマルチメディアポータル

筆者は、政策研究大学院大学の増山教授が主導する国会のマルチメディアポータルのプロジェクトに協力してきた（https://gclip1.grips.ac.jp/video/）。このシステムには、衆議院の会議録作成システムのために開発した音声認識システムが用いられている[4]。深層学習モデルの導入により、現在では 95%以上の文字正解率を実現している。このポータルでは、音声認識システムにより、すべての会議の映像の字幕を即日で生成する。公式の会議録が公開されると、前節で述べた方法により、字幕をそのテキストに差し替える。その際に、発言者の氏名も会議録に基づいて挿入される。 このシステムでは、キーワードや発言者に基づいて、映像中の該当する区間を検索し、ピンポイントで再生することができる。

さらに、地方議会向けに同様のポータルを構築している（https://localassembly-video.jp/）。 地方議会の YouTube チャンネルの URL を入力すると、音声認識によって生成された字幕付きの検索可能な動画のコレクションが用意される。地方議会の収録環境は国会ほど調整されていない場合もあり、また地方特有の固有名詞や方言も多いため、音声認識の正解率は 70〜90%程度に低下する。しかし、審議映像と字幕テキストの対応付けを行い、キーワード検索を実現することは可能である。これにより、誰でも簡単に議会の審議映像の検索が可能になる。

### 結論

審議映像の配信は、地方議会でも一般的になってきているが、多くの場合、字幕付与は行われていない。録画された映像のオンデマンド配信の場合、公式の会議録を使用する字幕付与が簡単な解決策であり、音声認識技術を使用することで、審議映像と会議録テキストの対応づけを行うことができる。

### 参考文献

[1] E. Voutilainen and R. Kuronen. Text Alternatives for Video Recordings in the Parliaments of Europe, 2024.

[2] T. Kawahara. Quantitative analysis of editing in transcription process in Japanese and European Parliaments and its diachronic changes. In ParlaCLARIN IV Workshop, pp.66-69, 2024. (https://aclanthology.org/2024.parlaclarin-1.10/)

[3] T. Korhonen, H. Kotze, and J. Tyrkkö eds. Exploring language and society with big data: Parliamentary discourse across time and space. John Benjamins, 2023. (https://doi.org/10.1075/scl.111)

[4] T. Kawahara, S. Ueno, and M. Morikawa. Transcription system using automatic speech recognition in the Japanese Parliament. The Journal of Professional Reporting and Transcription (Tiro), No.1, 2020.


国会審議映像検索システム ｜ 比較議会情報プロジェクト ｜ 政策研究大学院大学