



AIを活用した番組制作とユニバーサルサービス

日本放送協会 放送技術研究所 スマートプロダクション研究部

		部長	やまのうち ゆうこ 山内 結子	副部長	かねこ ひろゆき 金子 浩之
上級研究員	いまい あつし 今井 篤	上級研究員	さとう しょうえい 佐藤 庄衛	上級研究員	ことう じゅん 後藤 淳
		上級研究員	もちづき たかひろ 望月 貴裕	上級研究員	かわい よしひこ 河合 吉彦
主任研究員	みしま たけし 三島 剛	主任研究員	くまの ただし 熊野 正	主任研究員	うめだ しゅういち 梅田 修一

1. はじめに

NHKでは、新たな番組演出や迅速な報道、業務改革の実現に向けて、AI (Artificial Intelligence) 技術を活用する動きが活発化している。映像や音声、原稿などの膨大なデータや、蓄積されたノウハウを基にモデルを学習し、目的とするシステムに組み込んで利用することで、従来の作業を省力化し、記者やディレクターがより創造的な活動に注力する手助けとなる。

放送技術研究所 (以下、技研) では、画像解析や音声認識、テキストビッグデータ解析など、番組制作を支援するインテリジェント番組制作技術と、障害者や高齢者、外国人などに分かりやすく情報をお届けするユニバーサルサービス技術とを集約し、これらを「スマートプロダクション」と呼んで研究開発を推進している。

量に取得できるようになった。また、ソーシャルメディアの利用が一般化するにつれて、事件や事故などの第一報や、社会のトレンド情報などがTwitterをはじめとするソーシャルメディアから取得できるようになってきている。さらに、自治体等が公開しているセンサー情報などのオープンデータを活用して番組で取り上げることもできるようになってきた。このような大量の映像・音声素材から番組に必要な要素を取り出したり、ソーシャルメディアから番組に役立つ情報を取得したりすることは、番組制作者にとって非常に負荷の高い作業である。また、制作された番組などのコンテンツを外国人や聴覚・視覚障害者の方々を含む、全ての人に届けるためには、受け取る人の視聴環境に合わせたコンテンツの形式に変換する必要がある。

そこで、AI技術を活用してコンテンツ制作を支援するインテリジェント番組制作と、障害者や高齢者、外国人など全ての人にやさしい放送を届けるユニバーサルサービスに向けた研究開発を放送現場と連携して進めている (図1)。次章以降で、各要素技術におけるAI活用の事例を中心に、

2. スマートプロダクション

近年、放送局では、素材伝送回線の高速化や、記録メディアの大容量化に伴い、番組制作のための映像や音声を大



図1. スマートプロダクション

スマートプロダクションの研究開発状況と今後の取組みを紹介する。

3. テキストビッグデータ解析

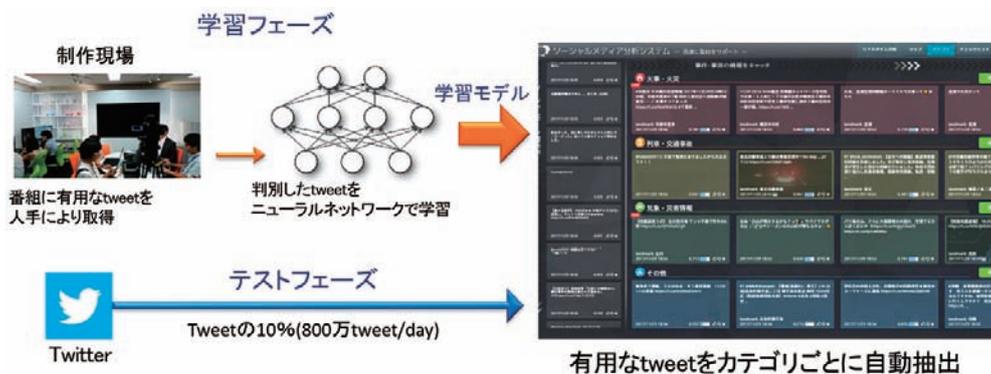
近年、Twitterなどのソーシャルメディアの利用が一般化するにつれ、個人がソーシャルメディアに投稿する情報は、既存メディアにとって不可欠な情報源となってきた。スマートフォンやタブレットをはじめとする高性能モバイルデバイスを利用するユーザーが爆発的に増えたことにより、個人が遭遇した事故や事件などをすぐにソーシャルメディアに投稿することが容易になっているためである。実際に、制作現場ではソーシャルメディアへの投稿を確認し、報道につながる取組みを実施している。例えば、2016年に起きた新幹線の放火事件では、その場に居合わせた乗客のソーシャルメディアへの投稿が放送の第一報として利用された。しかしながら、大量の投稿の中からニュース性のある投稿を人手で特定することは容易ではない。そのため、ソーシャ

ルメディアへの投稿をリアルタイムに自動で分析し、有益な情報のみを取得・提示する技術が必要とされている。

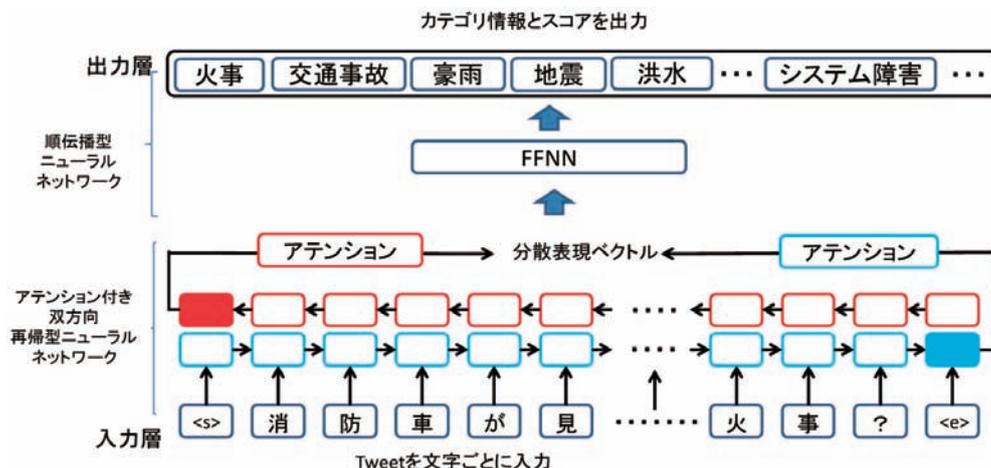
3.1 ソーシャルメディア分析システム

ソーシャルメディアのなかでも、Twitterへの投稿 (Tweet) を一種のセンサーのように活用して、対象の出来事を取得する多くの研究がこれまでも行われている。当所においても、制作現場のニーズに応えるため、ニュース価値のあるTweetを自動で特定するソーシャルメディア分析システム (Social Media Analysis System : SMAS) の開発を進めている (図2)。SMASでは、現場において人手でニュース性があると判断したTweetを教師データとして、ニューラルネットワーク (NN) を用いて学習している。これにより、ソーシャルメディアの大量の投稿の中から、ニュース価値のあるTweetを判別できる。

SMASでは、ニュース性判定の学習に、ニューラルネットワークを用いている (図3)。動作としては、まず、文章のよ



■ 図2. 開発中のソーシャルメディア分析システムの概要



■ 図3. システムで利用するニューラルネットワークの概要



うに連続する時系列のデータを取り扱うことができるニューラルネットワーク（再帰型ニューラルネットワーク：RNN）にTweetを文字単位で入力し、注目する表現に重みを付与する処理（アテンション機構）を行う。双方向に処理されたRNNのそれぞれの出力を結合することで、計400次元の分散表現ベクトルを得る。最後に、このベクトルを一般的なニューラルネットワーク（順伝播型ニューラルネットワーク：FFNN）に入力し、ニュースに有用なTweetを判別する。この出力をニュースカテゴリ数にすることで、ニュースに役立つ情報かどうかだけでなく、火事・火災、交通事故、豪雨・台風などニュースのカテゴリ情報も同時に判定することができる。

今後、開発しているシステムを制作現場で実際に利用することで、制作者がシステムの結果を取捨選択する履歴を基に学習を繰り返し、システムの性能を改善するメカニズムも取り入れていく予定である。

SMASの出力結果には、火事や火災などの表現が含まれていても、番組制作には利用できない場合がある。例えば、悪意のあるユーザからのデマ情報、いわゆるフェイクニュースである場合や、映画やテレビ、ゲームなどの仮想世界での出来事である場合がある。また、実際の事件や事故に関するTweetであっても、メディアが既に報道した記事であったり、その参照である場合も多い。そのため、仮想世界での出来事と現実の出来事を区別するための手法や、既に報道されている情報を特定する手法の開発を進めている。今後は、制作現場と連携して、フェイクニュースの特定を支援する技術の研究開発にも取り組む予定である。

3.2 原稿作成支援システム

放送局では、日々、様々な機関から幅広い情報を取得して、情報発信に役立っている。例えば、台風や地震などの災害の際、公共機関や地方自治体が提供している観測データなどである。このような大量のデータ、いわゆるビッグデータを人手で常時監視し、その結果を迅速に判断して放送のための原稿を正確に制作することは大変な作業である。そのため、緊急報道時や夜間などに、迅速で正確なコンテンツの制作を支援する技術は重要である。

放送局では、制作者が作成した原稿をデータベースに蓄積している。この原稿データベースは、新たな原稿を記述する際の参考としたり、過去の事実を確認するなど、放送局にとって重要な資産である。言語処理の研究にとっても、翻訳や音声認識の言語モデルなど多様な用途に利用できる言語資源である。

そこで、公的機関から取得した水位情報と原稿データベースを利用して、状況にあった原稿案を自動で作成する支援システムを開発している。図4に開発している原稿作成支援システムの概略を示す。システムでは、まず、過去に取得した水位情報と原稿データベースの対象原稿を対応させて、河川名、警戒水位名などのタグ付きコーパスを作成する。このコーパスをRNNによる固有表現抽出手法により学習し、得られた学習モデルにより過去の原稿中のタグの表現を自動で特定し、テンプレートを作成する。原稿から自動作成されたテンプレートとリアルタイムに取得する水位データを利用して原稿案の作成を行う。

現在、降雨時期に合わせて、原稿作成支援システムのプロトタイプを制作現場で試用しているところである。今後



図4. 原稿作成支援システムの概要

は、現場から要望に応じて、河川の水位以外のドメインについてもシステムを拡張していく予定である。

4. 画像解析

4.1 顔認識技術

映像アーカイブスや各放送局に蓄えられた映像素材は番組制作者にとって貴重な資料である。再利用の際に所望のシーンを迅速に見つけ出すためには、各シーンにキーワードを自動で付与する技術が不可欠である。

番組制作現場では、特定の人物が映る場面を素早く検索したいという要望が多い。そこで、映像素材に「人名キーワード」を自動で付与するために、映像中の顔が誰であるかを特定する「顔認識技術」の研究を進めている。

顔認識技術は、映像中の顔を見つける「顔検出」技術及び、その顔が誰であるかを判定する「顔識別」技術で構成される(図5)。テレビ映像は、照明条件、出演者の顔の向き、表情、鮮明度などが変動するため、顔の特徴点(目じりや唇の端点など)を手がかりとする従来手法では、顔の検出漏れや識別誤りが生じやすい。そこで我々はまず、目

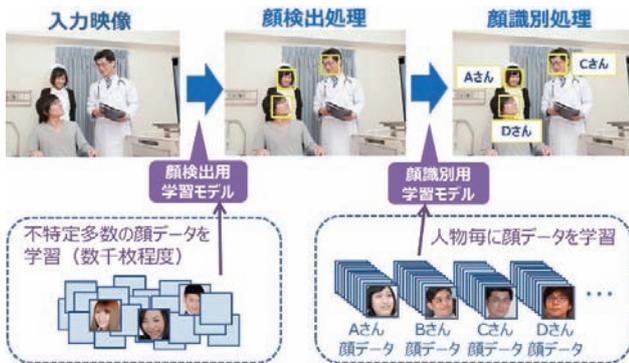
鼻の相対的な位置関係を考慮した画像特徴とカスケード型の決定木構造を持つ判定機の適用により、テレビ映像に対して高精度かつ高速な顔検出技術を開発した。さらに、顔認識技術についても、特徴点に依存しない領域ベースの特徴量(様々なサイズに分割された領域ごとに算出した特徴量)の導入及び識別モデルの改良により、テレビ映像に対する識別精度の向上を実現した。図6に処理例を示す。

この技術が、ある会社に設置された約100台のカメラによる撮影映像を題材とした番組の制作に活用された。特定人物が登場するシーンを顔認識技術で自動抽出する仕組みの実現により、番組映像を編集する作業コストの削減に大きく貢献した。

4.2 情景文字認識技術

映像に映っている現場の地名や建物名も、撮影現場を特定する情報として、映像を検索するための有用なキーワードとなる。そこで、映像に映りこんだ看板や標識などに書かれた文字列(情景文字)を検出して読み取る「情景文字列認識技術」の研究を進めている。

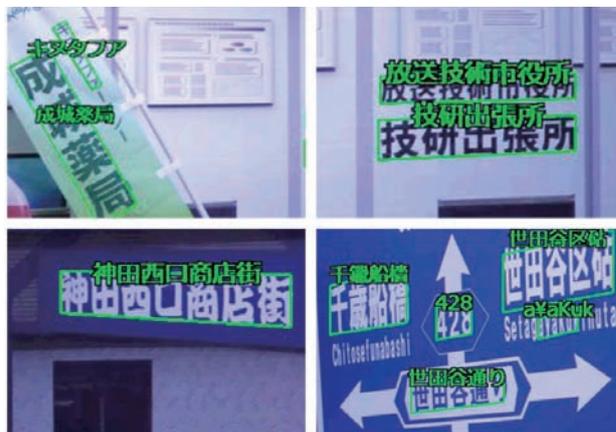
テレビ映像における情景文字列は、撮影方向の影響により、回転や変形が生じているケースが多い。さらに、大部分の文字は日本語であり、一部のひらがなや漢字が「パーツ」に分かれているため、正確な認識が難しい。そこで我々は、文字を任意に回転させた際のアスペクト比とパーツ配置を考慮した画像特徴及び文字の並ぶ方向を考慮した探索手法の導入により、回転や変形が生じた日本語情景文字列に対する認識精度の向上を実現した。さらに、大量の映像に対する処理を想定して、高速化にも取り組んだ。様々なパーツの組合せに対して「文字かどうか」を判定する処理



■図5. 顔認識技術の概要



■図6. 顔認識技術による処理例



■図7. 情景文字列認識技術による処理例



の効率化により、毎秒10フレームでの認識速度を実現した。図7に本技術による処理例を示す。

4.3 映像要約技術

番組の予告映像やダイジェスト映像は、番組をホームページなどで簡易に紹介するための重要なコンテンツである。我々は、それらのコンテンツ制作を支援すべく、番組映像から主要な部分をピックアップした「要約映像」を自動生成する技術の研究を進めている。本技術は、画像解析に基づいた登場人物、テロップ及びカメラワークなどの情報だけでなく、SNS解析による視聴者の反響も考慮した映像要約を可能とする。さらに、利用者（番組制作者）が上述の各情報の重み配分を自由に設定することで、「SNSで盛り上がった場面を重視」「顔のアップが多め」「カメラのズームインに注目」など、様々なパターンの要約映像を自動生成することができる。本技術の概要を図8に示す。今後、生成された要約結果の有効性について評価実験を行いながら、より精度の高い要約手法の開発に取り組む。

また、本技術を応用して、大量の投稿動画から要約映像を自動で生成するシステムを開発し、放送番組で活用された。一般の方、アスリート及びタレントがスマートフォンで撮影した約1400本の投稿動画の中から、顔の検出技術、建物や人ごみを認識する技術などを用いて映像を自動で選択し、それらをバランスよく並べて要約映像を生成する仕組みである。本システムにより2分30秒及び5分の要約映像



■図8. 映像要約技術の概要

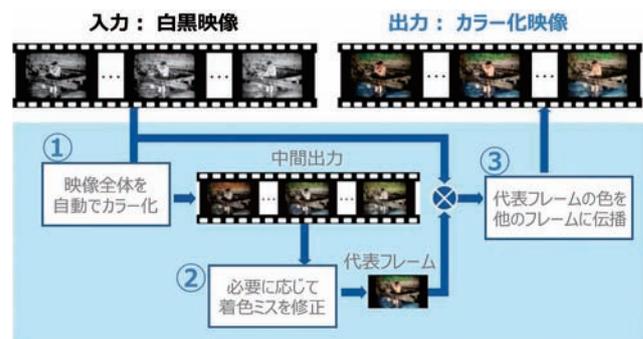
を合計6本自動生成し、番組において「AIが編集した映像」として紹介された。

4.4 白黒映像の自動カラー化技術

歴史番組やドキュメンタリーなどで活用する過去の白黒映像をカラーにできれば、当時の様子をより詳しく表現することが可能となる。しかし、従来の手作業によるカラー化では、映像を1枚ずつ人手で着色するため、多くの作業時間が必要となる課題があった。そこで、技研ではAIを活用した白黒映像の自動カラー化技術を開発した。AIの特徴を生かして、短時間で効果的にカラー化することが可能となった。

カラー化処理では、あらかじめ深層学習ニューラルネットワーク(DNN)に空や山などの色を覚えこませる必要がある。そこで、NHKアーカイブスに保存されている過去の番組など、合計約2万番組分の映像を使い、様々な物の色を学習させた。これにより、人手で数日かかる作業が、30秒から5分程度で行えるようになった(図9①)。しかし、DNNがいつも正しい色をつけられるとは限らない。そこで、映像中の特定の領域にどの色をつけるかを番組制作者が指定するだけで、容易に修正できる仕組みを開発した(図9②)。さらに、シーンを代表する1枚(フレーム)でつけた色の情報を、その前後のフレームに伝えることで、シーン全体を同じ色合いに調整する仕組みも開発した。これにより、全体で色のぶれの少ないカラー映像を作成できる(図9③)。

本技術を用いた自動カラー化システムが番組で活用された。海外で起きたある戦争を題材としたドキュメンタリー番組の制作において、当時の白黒フィルム映像をカラー化する作業コストの大幅な削減を実現し、制作現場から好評を得た。現在、さらなる番組での利用依頼を受けており、引き続き対応していく。



■図9. 自動カラー化技術の概要

■ 情報番組へのリアルタイム字幕付与



■ 取材映像の迅速な「書き起こし」



■ 図10. 音声認識技術の放送応用

5. 音声認識

音声認識技術の放送応用として、生放送番組への字幕付与と取材映像の書き起こしの支援を主たる目的として研究開発を進めている。

5.1 字幕制作システム(生放送への字幕付与)

字幕放送は、主に聴覚障害者や高齢者にテレビ番組の音声文字で伝えるサービスである。さらに、一般の方々にも公共交通機関の中や騒音のある場面での番組視聴に利用されている。人手をかけて事前に字幕を用意できる事前収録番組に対し、生放送番組では字幕を制作する手段が課題となる。音声認識技術は、このような生放送番組で迅速かつ正確に字幕を付与するために開発されてきた。

音声認識の研究開発の進展により、様々な種類の番組の認識精度が向上してきたが、100%正確な認識結果は得られないため、正確な字幕とするために認識結果は人手によって確認・修正されてから字幕として放送されている。人がいち早く認識結果を確認・修正できるように、発話の文末を待たずに、認識単語を逐次確定していくのが字幕制作用の音声認識の特徴である。

現在、定時の短いニュース番組では、番組音声を直接認識して字幕を付与している。一方、相撲中継や複数の話者が自由に発話する情報番組では、番組音声を直接認識しても、人手で誤りを修正可能な十分な認識精度が得られない。このような番組では、字幕専用のアナウンサーが番組の音声を復唱したものを認識して、誤りを修正可能な認識精度を確保している。さらに、人手が足りずに認識誤りを修正する手間をかけられない地域局発の生放送番組への字幕付与を目的として、事前に用意したアナウンサーの読

み原稿を利用した字幕送出技術も開発した。この技術は、番組音声の認識結果と読み原稿を比較して、どの文章が読み上げられたのかを推定して、原稿の該当文章を字幕として放送する仕組みである。

現在、発話の明瞭性や背景雑音の課題により復唱を要している番組の音声を直接認識するための研究開発も進められており、DNNの導入により認識精度は実用化にあと一歩というレベルに達してきている。さらに、発音辞書どおりに発音されないあいまいな言葉を認識するために、発音変形を推定する方法や、発音辞書を用いない音声認識手法(End-to-End音声認識)も開発している。また、多様な話題を扱う番組音声を認識するために、番組の話題と、その話題で出やすい単語や単語の連鎖を同時に推定するDNNの研究開発も進められている。これらの開発の結果、地域向けの情報番組の認識実験では、単語の誤認識率を35%から6.9%に削減することができた。

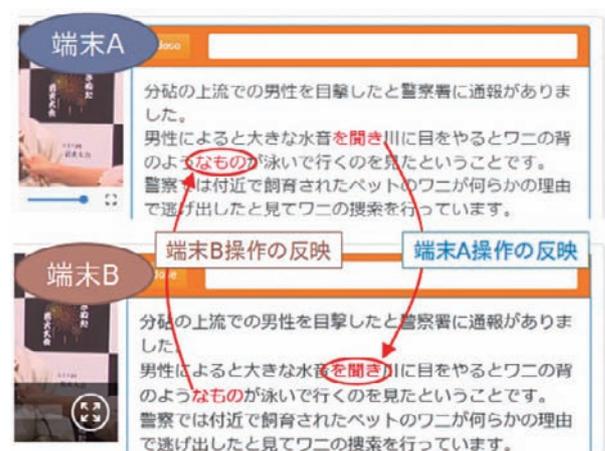
5.2 取材映像の書き起こし支援

放送局をはじめとする報道機関は、正確な情報をいち早く視聴者に届けることを使命としている。この情報の源となるのが取材映像であり、番組制作者は長時間にわたる取材映像を適切に編集して分かりやすく伝えている。この番組制作の過程では、取材内容を一覧して事実を確認するため、取材した映像に収められた発話の内容を文字起こした書き起こしが不可欠である。一方、この書き起こし作業は多くの労力を要し、番組制作の効率性や迅速性の妨げになっているため、この課題を解決する書き起こし支援システムが必要とされている。

そこで、音声認識を用いた書き起こし制作支援システム



■ 図11. 書き起こしインターフェース



■ 図12. 修正状況の共有

と、認識結果と映像・音声に対応づけて効率よく認識誤りを修正して書き起こしを完成させるインターフェース（書き起こしインターフェース）を開発した。

番組制作者は取材映像の音声を一字一句全て書き起こしてはいない。実際に取材に立ち会っている時点で、取材内容の概要は把握できているため、編集や事実関係検証のために重点的に書き起こすべき部分を理解している。そこで、インターフェースの開発にあたっては、素早く目的の取材コメントにアクセスして再生できるように設計した。

このインターフェースでは、映像の変換点や無音区間を手掛かりに取材映像を自動的に項目に分割し、それぞれの項目を特徴付け、話題の推移を把握するためのキーワードを自動付与した。このインターフェースは、まず、取材動画の該当部分の映像のサムネイルとこのキーワードを提示する。利用者はこれらの情報から書き起こし対象部分を特定して、クリックすることにより、認識結果確認と認識誤り修正画面へと効率よく遷移できる。

番組制作者はここで、認識結果の確認と認識誤りの修正の2つの作業を同時に行わなければならない。音声区間の再生と停止を繰り返しながら、認識誤り単語を修正する。提案するインターフェースでは、映像・音声の再生・停止する動作と文字を編集する作業を連動させたことにより、認識誤りを少ない操作で修正できるようになった。具体的には、認識結果の単語をクリックするとその単語が発話された時刻から映像・音声再生され、カーソルの移動や文字列の選択など認識誤り単語を修正するためのアクションがあると再生が停止する。

さらに、本システムはWebアプリケーションとして実装されており、ネットワークに接続していれば、自席からでも編

集内容を議論する場からでも書き起こしにアクセス可能である。Webアプリケーションとした場合には、複数の人が同時に同じ取材映像にアクセスすることが想定されるほか、複数人が同時に認識結果を修正する場合も想定される。本システムでは、複数の端末で修正状況をリアルタイムに共有できるため、複数人で協調して書き起こしを制作できる。記者会見発表など書き起こしが迅速に必要となる場合に有用な機能である。

書き起こし支援システムには、XD-CAMなどで取材してきた映像を、番組制作者がシステムにアップロードして書き起こしを制作するオフラインシステムと、様々な経路で伝送されてくる伝送素材を常時書き起こしているリアルタイムシステムがある。これらのシステムを放送現場から利用できるように設置し、運用実験を進めながら、システムの利用形態を調査するとともに、利用者からの要望も受けて改善を継続している。

書き起こし制作システムに用いられている音声認識は字幕制作用の音声認識が基になっており、現状では、放送音声と同等に明瞭な音声は精度よく認識できるが、取材映像には明瞭度が低く結果的に放送に用いられない部分も多く含まれる。事実確認や正確性の向上には、こうした部分の書き起こしも必要となるため、今後も明瞭度の低い音声を高精度に認識する技術の研究開発に取り組んでいく。

6. 音声ガイド

テレビのスポーツ生中継を対象に、視聴者が試合の状況を「聴いて」理解することができるように、各競技に付与されるリアルタイムデータから音声による解説「音声ガイド」を自動生成する技術の研究を進めている。この研究の目標

は、視覚障害者がテレビの中継番組をより楽しむことができるようにすること、またインターネットで配信される全ての競技映像について、音声による解説を自動制作し提供できるようにすることである。

6.1 音声ガイドの必要性

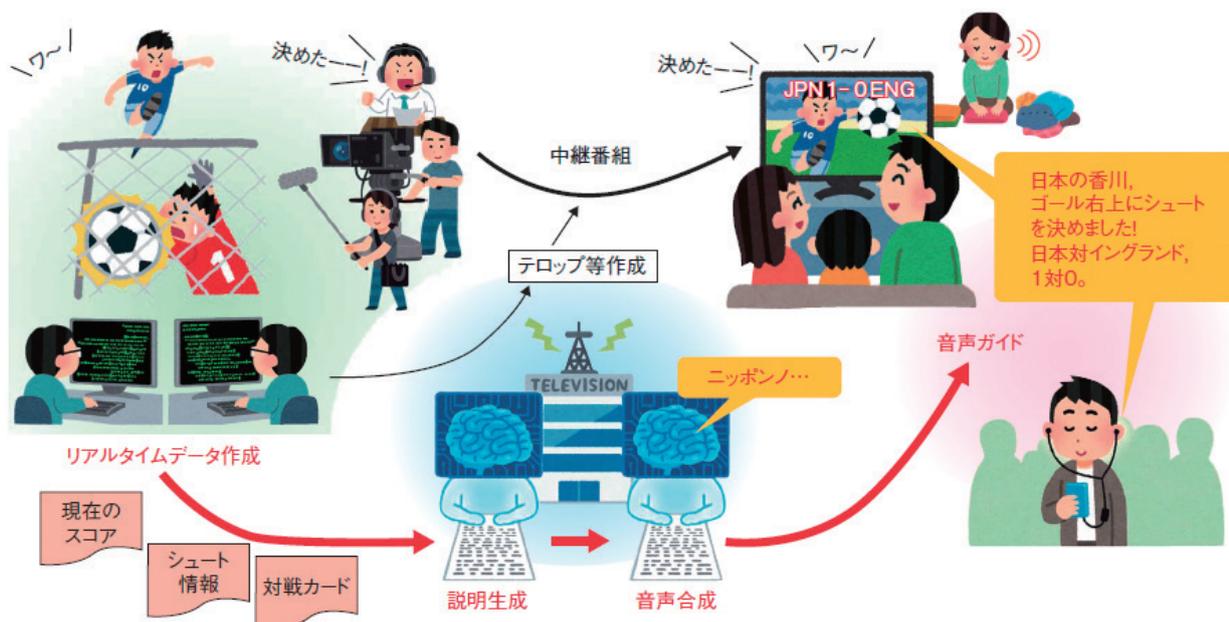
テレビはラジオと異なり映像から得られる情報は意図して発話されないことが多く、音声のみで番組を楽しむことが困難な場合が少なくない。テレビの高精細・大画面化に伴い、テロップなどの文字で伝達される情報量も増加しており、テレビの音声のみから番組の内容を把握することがますます難しくなっている。視覚に障害を持つ方々が家族と同じテレビ番組を楽しむ、あるいは家事や運転をしながら番組を楽しむといったニーズに応えるためには、映像に含まれる情報を音声で補完するサービスが必要である。放送局では、人の表情など音声だけでは伝わらない情報を音声で説明し、これを「解説放送」としてテレビの副音声で提供している。しかし、主音声に被らないようにすることが求められているため、ドラマなど限られたジャンルの収録番組にしか付与されていない。解説放送のための台本を制作したり、専用のキャストを用意するなど、コスト高も問題になっている。その結果、NHKが2016年度の全放送番組のうち解説放送サービスを提供した番組の割合は、総合で10.1%、Eテレで14.5%にとどまっている。

6.2 音声ガイド自動生成システム

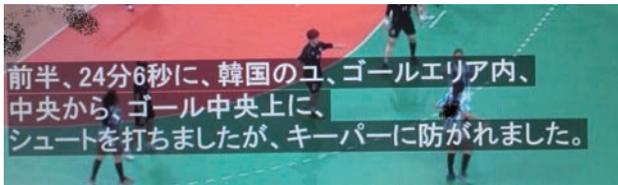
テレビのスポーツ中継の場合、アナウンサーの実況内容は「得点が入った」といったイベント情報（ある事象が発生したことを表す情報）が多く、得点や順位といった画面に常時表示されている情報は言及されないことが多いため、画面を見ずに音声のみを聴いていると試合の進行が把握しにくくなる場合がある。しかし、生放送番組に対する解説放送の提供は、技術的にもコスト的にも容易ではない。

現在、多くのスポーツ競技で、試合の得点状況や反則などのデータが、主催者や競技団体によってリアルタイムで作成・配信されている。これらのデータを用いて、放送局ではテロップやデータ放送など、画面に表示されるコンテンツを制作している。このデータから状況の変化に応じた説明文章を逐次生成し、これを合成音声で発話させることで音声ガイドのリアルタイム制作が可能となる。図13に制作フローを示す。

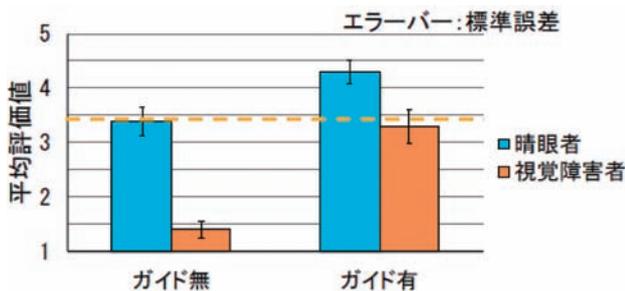
このシステムは、配信データを受信して説明内容を生成する説明生成部と、説明を音声化する音声合成部、現地映像（音声は会場音のみ）と、ガイド音声と字幕を重畳するガイド重畳部からなる。説明生成には、あらかじめ競技ごとに準備した発話テンプレート（「誰が」、「いつ」、「何をした」などを穴埋めする定型文面）を用いた。2016年のスポーツ国際大会では、現地から配信された映像とリアルタイム競技データを用いた動作検証を行い、約2000試合に音声ガイ



■ 図13. リアルタイム配信データを用いた音声ガイドの制作



■ 図14. 自動発話生成例



■ 図15. 試合の状況に関する内容把握の比較

ドを自動で付与した。合成音声の読み原稿はそのまま字幕としても利用できる。図14に自動発話生成例を示す。

今回作成した音声ガイドがどの程度試合進行の把握に役立つかを調べるため、晴眼者と視覚障害者、各10名を対象に、それぞれ音声ガイドの有無による内容把握度に関する主観評価実験を行った。図15に結果を示す。

対象競技は比較的情報量の多いハンドボールとし、映像と会場音のみのコンテンツを用いた。その結果、把握度1(非常に悪い)～5(非常に良い)の5段階評価の平均で、晴眼者が約3.4から4.3に、視覚障害者が約1.4から3.3に向上した。以上から、音声ガイドの一定の有効性が示された。

6.3 音声合成技術

音声情報を効果的に伝えるためには、伝えるべき内容に相応しい読み方が求められる。また、将来的に実況音声と共存させる際には、実況と音声ガイドの両方が聞き取りやすくなるような声質などについても検討する必要がある。自然で伝わる音声合成器を目指して、DNNを利用した新しい音声合成技術の研究を進めている。この技術は、NHKニュースのCGレポーターの声としても応用展開している。滑らかな発話をするだけでなく、目的に応じた「伝わる」発話を目指して、今後も研究を進めていく予定である。

テレビのスポーツ番組を対象にした、音声ガイドと自動実況をデータから発話させるサービスの可能性について示した。同様のサービスをスポーツ中継以外にも展開するため

に、引き続き各種データをリアルタイムに取得する技術と、音声合成技術の高度化について検討を進めていく。

7. 手話CG

日本語の代わりに手話を用いてろう者に分かりやすく情報を伝えるために、CGで手話を表現する技術の研究を進めている。手話CGを生成するために、あらかじめ手話の単語や句単位の動きをモーションデータとして記録している。モーションデータの記録には、手話者を赤外線カメラで撮影して手指や顔に付けられたマーカーの3次元位置を計測する、光学式モーションキャプチャーの手法を採用している。記録した単語や句のモーションデータを接続することで手話の文章を表現する動作を生成し、CGキャラクターを動かしてアニメーションを作成する。技研では、定型文の日本語文を自動合成する技術と、任意の日本語文を翻訳する技術の両面から手話CGの研究開発を進めている。

定型文の手話CGは、外部からのデータを自動解析してテンプレート方式でCGを生成する技術である。あらかじめ天気予報のために手話のテンプレートを用意し、気象庁から定期的に送られてくるXMLデータを用いて「天気」の内容や「気温」、「降水確率」の数値などのデータを差し替えることで、気象情報の手話CG映像を自動的に生成することが可能となる。ろう者による理解度実験で、手話表現が分かるかどうかを確認した実験の結果、正答率96%が得られ、本手法による手話情報提示の有効性が確かめられた。

2017年2月には、NHKオンラインに気象情報手話CG評価サイト (<https://www.nhk.or.jp/str1/sl-weather/>) を公開して、原則1日3回の自動更新で、関東7都県の気象情報を手話CGで提供している。

さらに、気象情報からスポーツ分野に应用を広げ、音声ガイドの研究と同様に競技中に配信されるリアルタイム競技データを基にして、試合中の競技状況やルールなどの手話CGを自動生成するシステムの開発を進めている(図16)。あらかじめ、選手名や点数などを差し替えられる手話CGのテンプレートを用意しておき、試合中は競技データをテンプレートに当てはめて手話CGをリアルタイムに生成する。ユーザは、スポーツ番組を見ながら、タブレットに映し出される手話CGを見ることで、現在の得点状況やプレイ中の選手のデータなど、補完的な情報が得られる。手話CGが再生される際にはタブレット側が振動することによって注意がうながされるため、テレビ画面に集中していても手話



■ 図16. スポーツ定型文手話CGシステム

CGの見落としを防ぐことも可能である。また、競技の進行に同期して再生される実況の手話だけでなく、ユーザが選択したタイミングでルール解説の手話CGを閲覧でき、競技の知識を深めることができる。定型文の手話CG生成システムには、事前にチェックすることで誤りのない手話を保証できることと、一旦自動合成を開始すれば、ほとんど人手をかけなくても運用できるメリットがある。

また、技研では任意の日本語を翻訳して手話CGを制作する技術について研究も進めている。手話は日本語と語彙や文法が異なる自然言語であるため、任意の話題の文章を日本語から手話に自動変換することは極めて難しい。そこで、翻訳技術の基礎として、NHKで放送している手話ニュースの日本語と手話表現の対訳を学習データとして整備する取り組みを進めるとともに、機械翻訳した結果に対して単語の語順や表現を手修正することで手話CGを制作できるシステムの開発も進めている。

今後、これらシステムの開発と聴覚障害者への評価実験

を進め、手話による情報提供の拡充に結びつけていく予定である。

8. おわりに

ここで紹介したテキストビッグデータ解析、画像解析、音声認識によって、放送局が社会の様々な情報や過去の番組アーカイブスを活用する際に、制作者は効率的で迅速かつ正確に番組で必要となる情報を取得でき、スムーズに番組制作できるようになると考えている。また、耳や目に障害のある方を含む全ての視聴者に、情報を正確に伝えるユニバーサルサービスの実現は、公共放送の重要な役割であり、重点的に取り組んでいることを紹介した。

NHKでは、技研における研究開発成果を導入する放送現場が近くにあるというメリットを生かし、放送現場との密接な連携及び現場導入を進め、最高水準の放送・サービスの実現を目指して、今後も引き続き研究開発を進めていく。