

## IV. 国内における字幕放送等の今後の展望等に関する調査

字幕放送等に関連する技術開発状況や通信における字幕付与等の国内における状況について、既存文献や Web サイトの情報をもとに文献調査を行うとともに、テレビ局、字幕番組等を制作している事業者、研究機関等を対象としたヒアリング調査を実施した。これらを踏まえ、国内における字幕放送等の今後の動向等を展望した。

### 1. 字幕放送等に係る技術開発状況

字幕放送等に係る新たな技術開発の状況について調査を行った。特に、生放送での字幕制作における省力化など現在の字幕放送等における課題に対応する解決策の観点から調査を行った。また、「視聴覚障害者向け放送普及行政の指針」における字幕付与の除外対象番組等を参考にして、字幕放送等の普及促進に当たっての課題を抽出し、運用面も含めてその解決策を整理した。

#### 1.1 字幕等付与の状況

平成 9 年 5 月、字幕番組・解説番組の放送努力義務の創設等を内容とする放送法等の一部改正がなされ、同年 11 月、郵政省（当時）が平成 19 年までに字幕付与可能な放送番組について字幕を付すことを目標とする「字幕放送普及行政の指針」を策定した。平成 18 年度実績では、字幕付与可能な放送時間に占める字幕放送時間の割合は NHK（総合）で 100%、民間放送キー5局平均で 77.8%となった。

さらに、平成 18 年 10 月から、今後の技術・サービスの進展を踏まえ、字幕放送等の拡充の推進に向けた施策の立案に資するため、「デジタル放送時代の視聴覚障害者向け放送に関する研究会」を開催し、同研究会における提言を踏まえ、19 年 10 月、9 年に策定した行政指針の字幕付与可能な放送番組の範囲を拡大するとともに、新たに解説放送に係る普及目標を追加した「視聴覚障害者向け放送普及行政の指針」を策定した。これを受け、NHK、在京キー5局、在阪準キー4局、テレビ大阪、在名広域4局、テレビ愛知が字幕拡充計画及び解説拡充計画を策定した。

各局における平成 22 年度の字幕放送等の実績と、平成 20～29 年の字幕放送の拡充計画は下記の通りである。

### 1.1.1 字幕放送等の実績（平成22年度）

平成22年度における各局の字幕放送等の実績（デジタル放送）は下記の通りである。

図表Ⅳ 1 1 各局の字幕放送等の実績（平成22年度）

#### <行政指針の普及目標の対象となる放送番組における字幕番組の割合>

	平成21年度	平成22年度
NHK（総合）	52.7%	62.2%
NHK（教育）	51.8%	52.5%
在京キー5局（注1）	89.0%	88.9%
在阪準キー4局（注2）	88.0%	85.6%
在名広域4局（注3）	75.4%	77.5%
全国の系列ローカル局（注4）	62.4%	61.2%

※2週間のサンプル週（平成22年5月31日（月）～6月6日（日）及び11月29日（月）～12月5日（日））における調査。

#### <総放送時間に占める字幕放送時間の割合>

	平成21年度	平成22年度
NHK（総合）	47.6%	56.2%
NHK（教育）	40.7%	42.6%
在京キー5局	43.9%	43.8%
在阪準キー4局	40.2%	41.4%
在名広域4局	37.9%	38.5%
全国の系列ローカル局	30.4%	31.4%

#### <行政指針の普及目標の対象となる放送番組における解説番組の割合（注5）>

	平成21年度	平成22年度
NHK（総合）		7.6%
NHK（教育）		11.2%
在京キー5局		1.4%
在阪準キー4局		1.0%
在名広域4局		1.4%
全国の系列ローカル局		0.8%

＜総放送時間に占める解説放送時間の割合＞

	平成 21 年度	平成 22 年度
NHK（総合）	5.7%	5.9%
NHK（教育）	10.0%	10.0%
在京キー5局	0.5%	0.6%
在阪準キー4局	0.4%	0.6%
在名広域4局	0.4%	0.6%
全国の系列ローカル局	0.4%	0.5%

＜総放送時間に占める手話放送時間の割合＞

	平成 21 年度	平成 22 年度
NHK（総合）	0.0%	0.1%
NHK（教育）	2.3%	2.4%
在京キー5局	0.1%	0.1%
在阪準キー4局	0.1%	0.1%
在名広域4局	0.1%	0.2%
全国の系列ローカル局	0.1%	0.1%

（注1）在京キー5局：日本テレビ放送網（株）、（株）TBSテレビ、（株）テレビ朝日、（株）フジテレビジョン、（株）テレビ東京

（注2）在阪準キー4局：（株）毎日放送、朝日放送（株）、讀賣テレビ放送（株）、関西テレビ放送（株）

（注3）在名広域4局：中部日本放送（株）、東海テレビ放送（株）、名古屋テレビ放送（株）、中京テレビ放送（株）

（注4）全国の系列ローカル局：在阪準キー4局及び在名広域4局を除く101社

（注5）行政指針の普及目標の対象となる放送番組における解説番組の割合については、平成22年度より新たに集計・公表を実施

資料：総務省「平成22年度の字幕放送等の実績」

([http://www.soumu.go.jp/menu\\_news/s-news/01ryutsu05\\_01000012.html](http://www.soumu.go.jp/menu_news/s-news/01ryutsu05_01000012.html))

別添より

生放送については、「字幕放送普及行政の指針」において技術的に字幕を付すことができない放送番組としてニュース・スポーツ中継等も含めて対象から除外されていたが、平成 19 年に策定された「視聴覚障害者向け放送普及行政の指針」にて字幕付与可能な放送番組の定義が拡大されたため、新たに複数人が同時に会話を行う場合以外の生放送番組が字幕付与可能な放送番組に含まれることとなった。

(参考) 行政指針における字幕付与対象に関する改正点

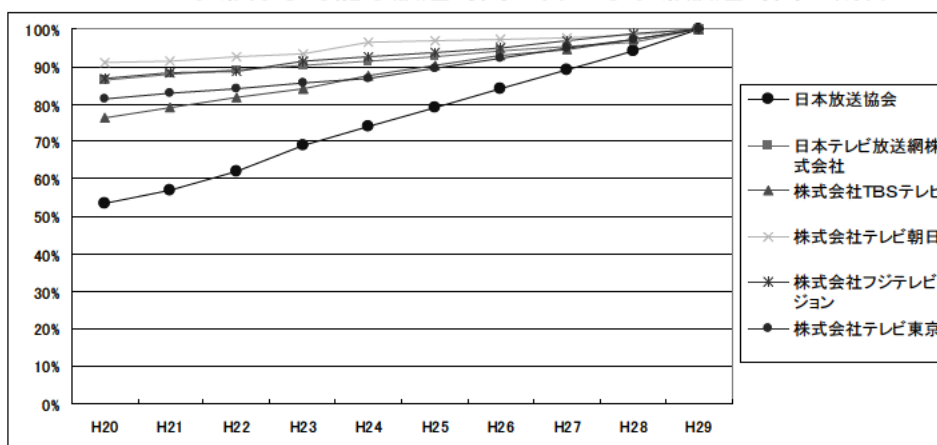
字幕放送普及行政の指針 (平成 9 年 11 月)	視聴覚障害者向け放送普及行政の指針 (平成 19 年 10 月)
対象：7時から24時までの放送番組 <除外対象番組> ・技術的に字幕を付すことができない放送番組（例 ニュース・スポーツ中継等の生番組） ・オープンキャプション、手話等により音声を説明している放送番組（例 字幕付映画、手話のニュース） ・外国語の番組 ・大部分が歌唱・器楽演奏の音楽番組 ・権利処理上の理由等により字幕を付すことができない放送番組 ・再放送番組	対象：7時から24時までの放送番組 <除外対象番組> ・技術的に字幕を付すことができない放送番組（例 現在のところ、複数人が同時に会話を行う生放送番組） ・外国語の番組 ・大部分が器楽演奏の音楽番組 ・権利処理上の理由等により字幕を付すことができない放送番組

1.1.2 字幕放送の拡充計画（平成 20～29 年度）

字幕付与可能な放送時間に占める字幕放送時間の割合、総放送時間に占める字幕放送時間の割合のそれぞれについて、NHK、在京キー局、在阪準キー4局、テレビ大阪、在名広域4局、テレビ愛知のそれぞれが、字幕放送の拡充計画を提出している。ここでは、例として NHK 及び在京キー局の平成 20～29 年度における字幕放送の拡充計画を示す。

図表IV 1 2 NHK及び在京キ 5局の字幕放送の拡充計画（平成20～29年度）

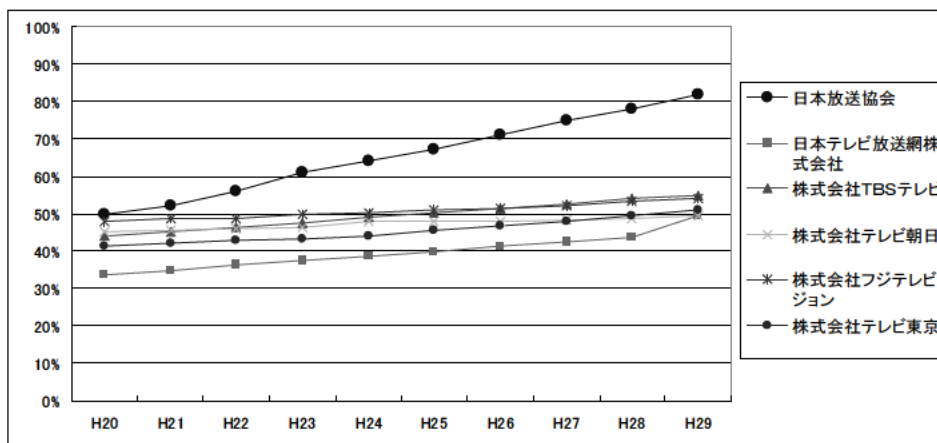
＜字幕付与可能な放送時間に占める字幕放送時間の割合＞



(注1) 各社の字幕付与可能な放送時間(週)は、NHK：7080分、日本テレビ：3788.5分、TBS：4392分、テレビ朝日：4388分、フジテレビ：4355分、テレビ東京：4989分（7時から24時までの17時間は、1週間で7140分）

(注2) 計画作成上、放送事業者の事情によらない突発的な事態が発生し、番組編成上の大きな変更があった場合は考慮していない。

＜総放送時間に占める字幕放送時間の割合＞



(注1) 平成22年度まで、アナログ放送とデジタル放送で計画値が異なる場合、デジタル放送の計画値を記載。

(注2) 計画作成上、放送事業者の事情によらない突発的な事態が発生し、番組編成上の大きな変更があった場合は考慮していない。

資料：総務省「平成22年度の字幕放送等の実績」

([http://www.soumu.go.jp/menu\\_news/s-news/01ryutsu05\\_01000012.html](http://www.soumu.go.jp/menu_news/s-news/01ryutsu05_01000012.html))

参考2をもとに作成

## 1.2 字幕放送等に係る新たな技術の開発状況

### 1.2.1 字幕放送

#### (1) 現状

字幕放送は、NHK（総合）で総放送時間の6割弱、NHK（教育）及び民間放送主要局では4割前後実施されている。

現在、NHK及び民間放送主要局では、収録番組の場合、納品が放送直前となり字幕を付ける時間的余裕がない場合等を除き、ほぼ全ての番組に字幕が付与されている。

一方、1.1.1で示した通り、「視聴覚障害者向け放送普及行政の指針」にて字幕付与可能な放送番組の定義が拡大され、例えば、複数人が同時に会話を行う生放送番組のように、技術的に字幕を付与することができないような場合以外の生放送番組については、字幕付与可能な放送番組の対象とされたことによって、生放送番組に対する字幕付与が新たな課題となっている。各局では、生放送番組への字幕付与に対応するに当たり、ニュースといった生放送番組が集中する時間帯や、早朝・深夜といった体制が組みづらい時間帯における番組には、字幕付与に対応できないケースが出てきている。

これらの現状を踏まえ、まずは生放送番組での字幕付与に当たり、現在主に採用されている4つの方式について下記のとおり整理を行った上で、字幕放送等に係る新たな技術開発について、それぞれの技術・システムの概要、特徴、課題等を整理する。

図表IV 1 3 生放送での字幕付与方法

方式	ダイレクト方式	リスピーク方式	テキストデータ 送出方式	リレー方式 (キーボード入力)
概要	発言者の発話を直接音声認識エンジンでテキスト化し、校正のうえ字幕化。	発言者の発話をうけ、リスピーカーが要約等を行ったうえで再度発話して音声認識エンジンでテキスト化。	放送用原稿をあらかじめ字幕化し、タイミングをみて送出する方式。送出自動化も実用化。	発言者の発話を聞き、入力者がキーボードでテキスト入力し、校正のうえ字幕化。速記をベースにした入力装置(高速入力用キーボード)が実用化。
利用技術	音声認識技術	音声認識技術	字幕送出	高速入力用キーボード(速記用)
必要人員	・校正作業員：1	・リスピーカー：1 ・校正作業員：1～2	・字幕送出(手動)：1	・入力者：4～6
メリット	・人員負担小	・音声認識向上(校正負担減)	・人員負担小(字幕送出1名)	・複数人が会話を行う場合であっても字幕付与可能
デメリット	・音声認識率が低い(校正の負担大) ・単語登録等の事前準備が必要	・リスピーカーが必要 ・単語登録等の事前準備が必要	・放送用原稿が必須	・人員負担大(入力者の確保) ・速記タイプの場合、入力者の育成
適用番組	ニュース、野球中継(NHK)	スポーツ番組、歌番組、情報・バラエティ番組(NHK)	ニュース(民間放送局)	ニュース：速記タイプ(NHK、民間放送局)、ニュース：キーボード(民間放送局)、歌謡/情報番組：キーボード(NHK)、スポーツ番組：キーボード(民間放送局)

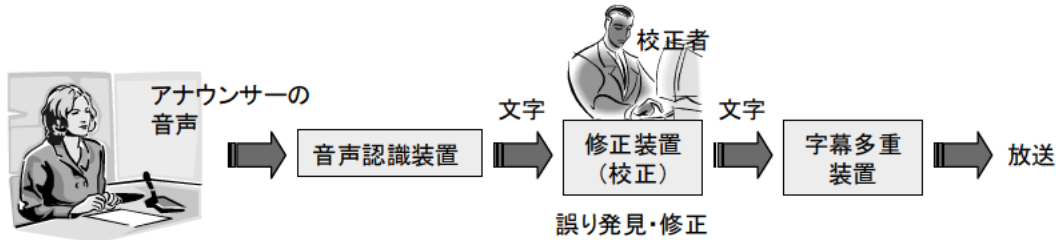
## (2) ダイレクト方式

### ①現状

#### <概要>

音声認識技術を利用し、アナウンサーの発話を直接音声認識エンジンによりテキスト化し、校正のうえ字幕化する方式。

図表IV 1 4 ダイレクト方式による生字幕制作



#### <必要人員>

校正作業者：1～4名（NHKの場合はチーフプロデューサー1名、技術者1名、校正者1名等の計4名程度）

#### <メリット>

- ・必要となるスタッフが少なくて済み、人員負担が小さい。
- ・原稿読み上げ部分については、一定の音声認識率（98%）が得られる。

#### <デメリット>

- ・音声認識装置において、単語登録等の事前準備が必要となる。
- ・原稿読み上げでなく自由発話部分については音声認識率が低い（校正の負荷が大きい）。ニュースでの街頭インタビューや中継は自由発話であるため、対応が困難。また、複数の話者がいる番組や雑音が多い場合には対応が困難。

#### <適用番組>

NHKでは、解説、アナウンスのみで構成され、それ以外の雑音が入らない番組がダイレクト方式の対象となっており、現状では米メジャーリーグ野球中継で採用されている。野球、相撲をはじめとするスポーツ番組は用語も限られており、音声認識率が高く、実用段階にある。

ニュースで利用された実績もあったが、中継など対応できないパートはリレー方式に頼らざるを得ないことから、現在、ニュース番組ではリレー方式（速記タイプ）に一本化されている。

東日本大震災時には、平成23年3月15日～18日の4日間に放送された一部の時間帯でダイレクト方式による字幕付与を試行した。



## ②技術開発例 1

### ○名称（開発主体）

ハイブリッド型字幕制作システム（NHK）。

### ○技術開発の目的

ダイレクト方式では原稿読み上げ部分に限定されていた字幕付与可能範囲の拡大。

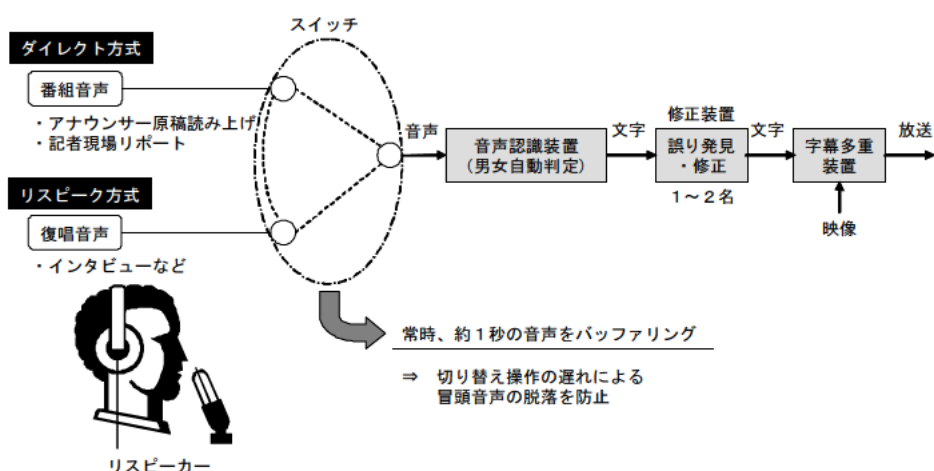
アナウンサーの原稿読み上げにはダイレクト方式を適用し、自由発話からなる街頭インタビューをはじめダイレクト方式では高い認識精度が得られない場合にリスピーク方式に切り替えることで対応する。

### ○技術・システムの概要

音声認識装置では、入力音声の波形をもとに発話区間を特定し、男女いずれの発話であるかを自動判別する。音声認識エンジンでは、電子化されたニュース原稿をもとに様々な単語の接続確率を事前に学習した言語モデルや、各単語の発音記号を収録した発音辞書等を利用して、発話内容に該当する文字列を探索し、認識結果を出力する。

音声認識装置が出力する単語列は、修正装置のタッチパネルモニターにリアルタイムで表示され、オペレーターが対応する音声を聞きながら確認する。誤認識の単語を発見すると、それを指でタッチして選択し、キーボード入力にて修正し、字幕となるテキストを送出する。

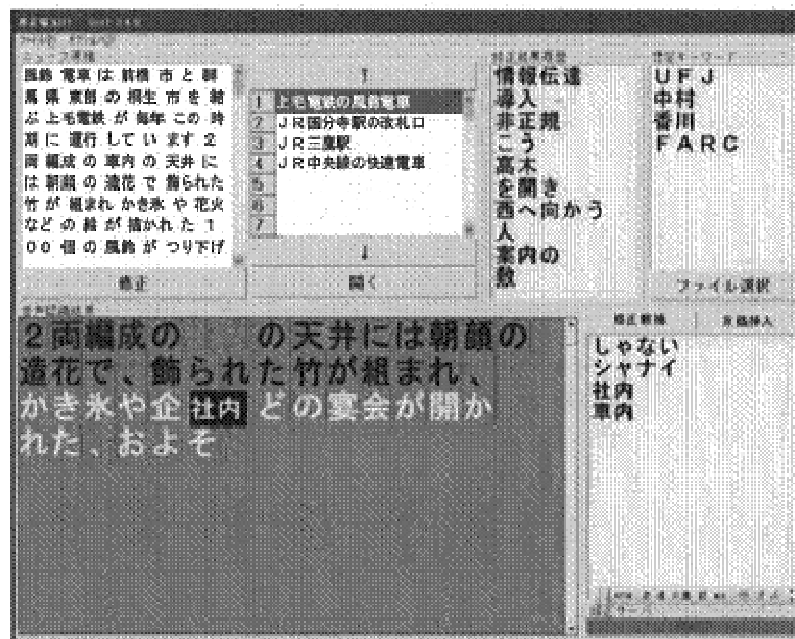
図表Ⅳ 15 ハイブリッド型字幕制作システムの概要



資料：本間真一（NHK放送技術研究所）「生字幕制作のための音声認識」

(<http://www.nhk.or.jp/str/publica/rd/rd122/pdf/P25-31.pdf>) より作成

図表IV-1-6 ハイブリッド型字幕制作システムの修正装置画面



資料：本間真一（NHK放送技術研究所）「生字幕制作のための音声認識」  
 (<http://www.nhk.or.jp/str/publica/rd/rd122/pdf/P25-31.pdf>) より

### ○技術・システムの特徴（従来技術との比較、効果）

ダイレクト方式とリスピーク方式を組み合わせることで字幕付与可能な範囲が広がり、音声認識装置を番組全体に適用できるようになることができる。また、入力音声の男女自動判別機能の導入により、従来自動判別機能がなかったために男女別に必要であった音声認識装置が1台に統合された。

修正装置は従来誤り発見用と誤り修正用の端末をセットで運用していたが、本システムでは装置の台数を想定される音声認識精度に応じて変更できるようになった。さらに音声認識率も向上<sup>1</sup>し、発話内容がある程度限定された記者の現場リポートもダイレクト方式で認識可能となった。また、ダイレクト方式のみで対応する場合に4名程度必要としていたオペレーターを、ハイブリッド方式で対応することによって1～2名に削減できた。

### ○課題

ニュース番組に出演する解説者がニュース解説を行う部分はリスピーク方式で

<sup>1</sup> スタジオのメインアナウンサーの原稿読み上げ部分の認識率は95%から98%に向上。音声認識率は（（発言単語数－修正単語数）÷発言単語数）で算出しており、認識率95%では6秒に1単語程度の誤認識に相当する。

対応するが、前述の原稿読み上げ部分をダイレクト方式で音声認識する場合よりも音声認識精度が低くなり、修正作業に負担がかかって遅延が生じるようになる。ニュースの場合には正確性が重視されるため、95%以上の認識率が求められるが、解説者がニュース解説を行う部分が増えると、この水準を下回ってしまうこともあるため、ハイブリッド方式による対応が難しくなる。

#### ○実用化の状況（実用化の目途）

平成 23 年度末に導入予定。

### ③技術開発例 2

#### ○名称（開発主体）

自由発話音声認識（NHK）。

#### ○技術開発の目的

自由発話部分の音声認識率向上及びリスピーク方式からダイレクト方式への移行。原稿・台本がなく、出演者が自由に発話する部分（自由発話部分）を多く含む番組については、従来リスピーク方式が適用されてきたが、ダイレクト方式により自由発話部分の音声を直接認識する精度を向上させ、字幕付与を行うことで、効率化に役立てる。

#### ○技術・システムの概要

発声があいまいとなる傾向の強い自由発話音声に対する誤認識の傾向を学習することで誤認識の識別能力を高める手法を導入し、認識精度を向上させる。ニュース用の言語モデルの学習では、大量の電子化されたニュース原稿が利用できるが、自由発話である話しことばの学習では、大量の電子化されたテキストは存在しないため、新たに書き起こし文を作成する必要があり、コストがかかる。そこで、大量の電子化された原稿に少量の自由発話の書き起こし文を適度に混合したテキストを作成・利用することで、自由発話の特徴をモデルに反映させる方法を検討している。

#### ○技術・システムの特徴（従来技術との比較、効果）

報道情報番組「クローズアップ現代」のゲスト対談部分を使用した評価では、研究着手時点での認識率 75.5%から、モデルの改善により認識率が 88.7%に向上した。

## ○課題

報道番組としては、まだ十分な認識率（報道番組において実用化可能なレベルは95%程度）に達しておらず（前述の通り現状の認識率は88.7%）、さらなる精度の向上が必要となっている。自由発話音声の不明瞭さ（早口等に見られる発声変形）や話しことばの対策による改善が進められている。

## ○実用化の状況（実用化の目途）

認識精度向上を目指した研究を実施中。

### (3) リスピーク方式

#### ①現状

##### <概要>

専任の話者（リスピーカー）がアナウンサーの発話を中心とする番組音声を聞き取り、アナウンサーの発話内容のうち、番組の映像を見れば理解できる部分については要約・省略を行い、かつ、アナウンサーの発話内容に含まれない拍手や歓声等については補足したものを発話する。音声認識技術を利用して音声認識エンジンでテキスト化する（現状の認識率95%）。民間放送局では、発話音声の認識結果を分割して複数の校正作業者に振り分けるシステムを自動化することで、各自の作業範囲を明確化し、作業者間でのコミュニケーションを不要にする作業分散化システムも実用化されている。

NHKでは、アナウンサーやリポーターのOB・OG等がリスピーカーを担当しており、それぞれスポーツ、情報といった得意分野を持っている。音声認識装置はリスピーカーごとに学習・適応させ、番組のジャンルに応じてコーパス（辞書）を作成する<sup>2</sup>ことで実用化が可能となる。

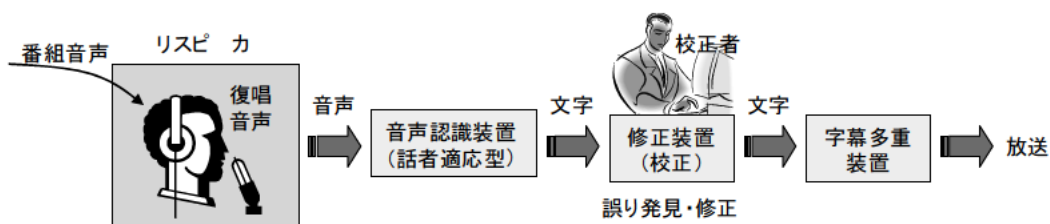
ワイドショー等の複数人が話す生放送の情報番組では、原稿や台本がなく、話題も変わっていくため、単語登録等の事前準備が十分にできず、ダイレクト方式による音声認識では対応が難しい。そこで、複数人が話す生放送の情報番組ではリスピーク方式での対応が進められている。NHKのアナウンサーOBがリスピーカーとなり、文字にした際にわかりやすい表現に言い換える。複数人の発話を全てリスピーカーがアレンジし、色分けされたボタンを押してから言い直すことに

---

<sup>2</sup> 例えばニュースの場合にはニュース原稿から収録語を取り込む。

よって、画面の字幕では発話者名とともに色分けされたテキストが表示される。

図表IV 17 リスピーク方式による生字幕制作



#### <必要人員>

リスピーカー：1名、校正作業員：1～2名。

#### <メリット>

- ・背景雑音の大きい番組、複数の出演者が話す番組にも対応できる。
- ・専任の話者に適応した音声認識装置を使うことで、音声認識率の向上（校正負担の軽減）が期待できる。
- ・発話にない拍手や歓声、状況等を補足できる。
- ・校正作業員の育成は1～2週間で可能であり、他の方式と比べて容易である。

#### <デメリット>

- ・要約、アドリブが入るため、番組音声を完全に再現するのが難しい。
- ・リスピーカーの人員負担。
- ・話者に合わせた音声認識を行うための音声登録、単語登録等の事前準備（2時間程度）が必要となる。

#### <適用番組>

NHKでは、歓声をはじめとする背景雑音、話しことば特有の不明瞭な発話等の原因により、番組音声を直接音声認識させることが困難な番組（スポーツ番組、歌番組等）を対象としている。音声認識装置に登録可能な語彙数を拡張（4万語→10万語）したことなどにより認識率が向上したため、適用可能な番組数が増えており、平成22年4月から「スタジオパークでこんにちは」での運用が行われるなど、近年では情報・バラエティ番組への適用も行っている。

また、一部民間放送局では、野球中継に利用している。

#### (4) テキストデータ送出方式

##### ①現状

###### <概要>

放送用原稿からアナウンサー向けの注釈といった不要な情報を削除し、字幕表示に適したテキストデータに加工しておき、運用スタッフが放送の進行状況をモニターしながらタイミングに合わせて順に送出させていく方式。送出を自動化するシステムも実用化されている。

NHK では、台本どおりに進行される番組に適用されている。また、他の方式と組み合わせて利用されることも多い（放送時に原稿と異なる内容になりそうな部分については他の入力方式等で対応）。

###### <必要人員>

字幕送出（手動）：1名 ※送出を自動化した場合には不要。

###### <メリット>

- ・運用スタッフは、字幕送出が手動の場合でも1名（自動化した場合は不要）となり、人員負担が小さい。
- ・放送前の準備作業の時間が30分程度と、他の方式と比べて短縮される。
- ・送出を自動化した場合には運用人員は不要となり、字幕表示の遅延もほとんどないことから、緊急放送の字幕付与にもあらかじめ準備可能な定型的表現であれば適用できる。

###### <デメリット>

- ・放送用原稿が必須で、話者が原稿どおりに発話しない場合には対応できない。
- ・放送用原稿に作成者の癖がある場合、それを直すための手間がかかる。

###### <適用番組>

ニュース番組をはじめ、原稿や台本等が固まっているものを対象としている。最近では、原爆式典で採用された。

##### ②技術開発例

###### ○名称（開発主体）

全自動リアルタイム字幕制作システム（NTT コミュニケーションズ）。

## ○技術開発の目的

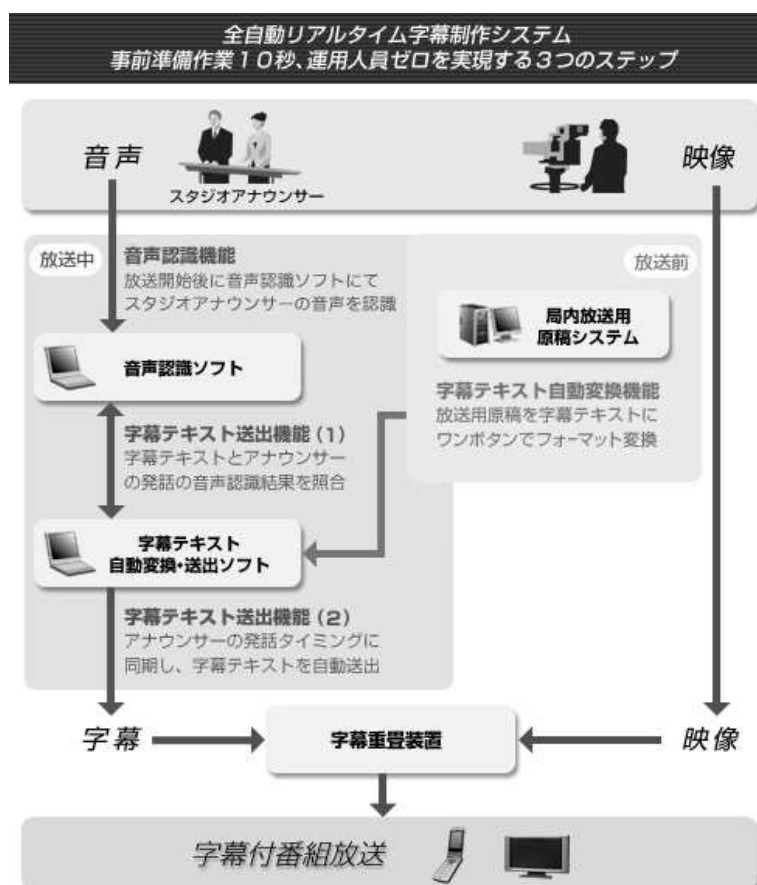
放送局における生放送番組に対する字幕付与作業の省力化。

## ○技術・システムの概要

放送用原稿は、自動変換機能によりワンクリックで自動的に字幕テキストに変換される。一方で、放送中のアナウンサーの発話は、NTT サイバースペース研究所が開発した音声認識技術（VoiceRex）により音声認識され、テキスト変換される。この技術は放送局との実験結果を反映して認識精度を高めており、アナウンサーが抑揚をつけて話した場合でも認識率90%台に達する。

最後に、アナウンサーの発話の音声認識結果と、事前に自動生成した字幕テキストを照合した上で、アナウンサーの発話のタイミングに合わせて字幕をリアルタイムに自動送出する。

図表Ⅳ 18 全自動リアルタイム字幕制作システムの概要



資料：NTTコミュニケーションズHP

(<http://www.ntt.com/ict/library/future/jimaku.html>) より

## ○技術・システムの特徴（従来技術との比較、効果）

放送前の事前準備としては字幕テキストへの自動変換に要する約 10 秒のみであり、運用スタッフが不要となる。また、字幕表示の遅延 1～2 秒と短く、ほぼリアルタイムで字幕が表示される。

## ○課題

字幕テキストの作成のために放送用原稿が必要となるため、原稿がない番組には適用できない。また、複数人が同時に発話した場合は音声認識が困難になり、字幕を修正する人員が必要となる。実際に導入する際には自動変換機能と音声認識機能のカスタマイズが必要となり、特に地方局にとってはそのコスト負担が大きい。

## ○実用化の状況（実用化の目途）

平成 20 年 1 月より商用化サービス開始（導入例はなし）。

## (5) リレ 方式（キ ボ ド入力）

### ①現状

#### <概要>

アナウンサーの発話を聞き取り、入力者がキーボードでテキスト入力したものを、校正者が修正した上で字幕化する。速記をベースにした入力装置（高速入力用キーボード）が実用化されており、この装置を使って専門事業者が高速入力を行う方法と、局内やグループ会社で通常のキーボードで入力する方法がある。スタッフや設備を全て局内に配置する必要はなく、遠隔地での対応も可能である。

#### <必要人員>

入力・校正・送出スタッフ：4～10 名。

#### <メリット>

- ・複数人が発話する場合や、放送用原稿・台本がない場合等でも字幕付与が可能であり、適用範囲が広い。
- ・制作（字幕入力）現場とテレビ局が離れていても字幕制作は可能であることから、地方局にも適用できる。

#### <デメリット>



- ・入力者が必要となり、音声認識を利用する方式に比べて人員負担が大きい。
- ・高速入力を行う方法では、専門スタッフの確保が必要となる。専門のスタッフは限られており、育成にも時間がかかるため、需要が集中する時間帯には人手不足となる。

#### <適用番組>

NHK では、報道系を高速入力タイプ、歌謡番組・情報番組等を通常キーボード入力タイプといったかたちで使い分けている。

民間放送局では、いずれの方法についても、ニュース、スポーツ、情報・バラエティ番組等、幅広いジャンルで適用されている。

#### ②技術開発例

##### ○名称（開発主体）

高速入力用キーボード「ステノワード」（スピードワープロ研究所）。

##### ○技術開発の目的

テレビ放送における発話内容の高速入力、リアルタイムでの字幕データ作成。

##### ○技術・システムの概要

テレビ放送における発話内容を高速でテキスト入力し、リアルタイムで字幕データを作成する。

母音 5 種と子音 5 種の 10 種のキーが横に並び、この 10 個のキーの組み合わせで日本語の高速入力ができる。

図表Ⅳ 19 高速入力用キーボード「ステノワード」

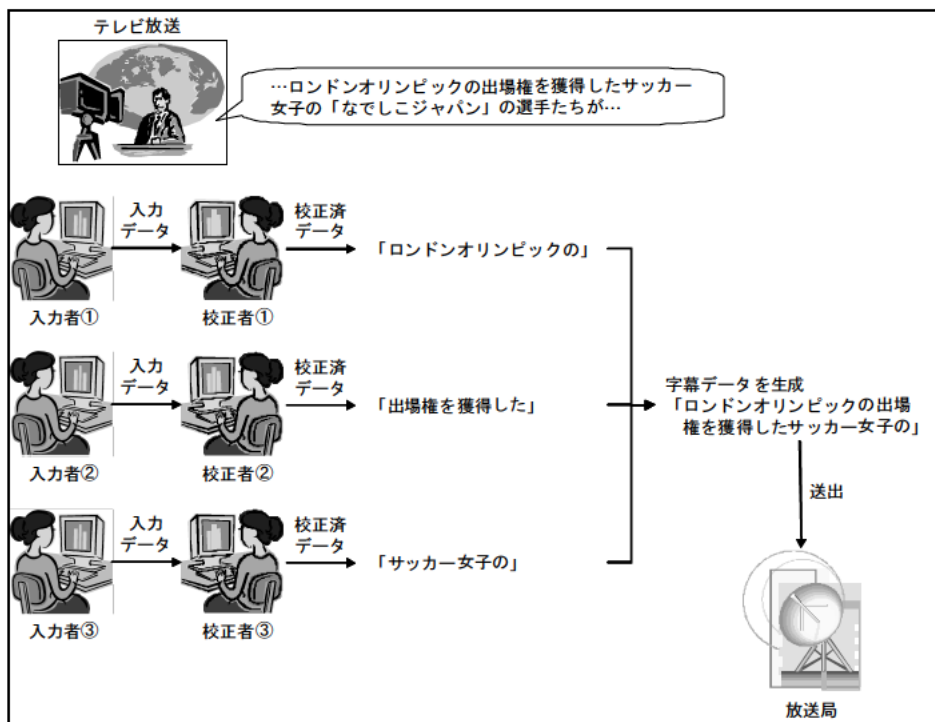


資料：スピードワープロ研究所HP（<http://www.speed-wp.co.jp/laboratory/>）より

キーボード入力後、校正した文字データを放送局に送出する。入力者 1 名と校正者 1 名の組み合わせを 3 組配置し、発話を区切りながら順に入力していく方式

で字幕を制作している。3組が並行して入力する際に抜けや重なりが出ないように、ランプの点灯、ヘッドホンからの音声出力（両側／片側）、表示画面による確認、及び入力者による先頭文字の発声等により制御している。

図表IV 1 10 高速入力システムの概要



### ○技術・システムの特徴（従来技術との比較、効果）

習熟すると1分間に300文字以上<sup>3</sup>と、通常のキーボードよりも高速で入力が可能となる（実際の放送における発話の速度は1分間に400文字程度。）。リアルタイムで字幕オペレーターが入力する方式は、事前の作業負担や精度、情報量といった観点から、他の方式と比べて最も実務的である。

字幕データは、放送での発話後1～2秒以内に送出されている。（実際の放送では一定間隔で表示されるようテレビ局側で調整。）

### ○課題

早朝から深夜まで様々な番組に字幕を付与するために、オペレーターのやりくりが課題となる。連続入力時間は1人あたり、1～1.5時間程度としており、長時間の番組や24時間放送の場合には、交代要員・チームを確保する必要がある。

また、オペレーターの育成にも時間がかかる。生放送のオペレーターとしては、

<sup>3</sup> 日本情報処理検定協会主催のパソコンスピード認定試験では、最高位の初段で10分間で1500文字以上（1分あたり150文字程度）の入力が要求される。

検定試験 1 級（320 字／分以上の入力ができる）が最低条件であり、そこまでの育成に 1～2 年程度を要する。さらに実地訓練が必要となるため、当初から現場に出るまでには 3～4 年かかる。放送用の原稿がない場合でも入力できるよう、オペレーターの国語能力も高めなければならず、そのために新聞の書き写しや国語辞典を使った全ページ書き取り学習も必要となる。

そのほか、放送時間帯が重なり、各局の需要が集中した場合には、それらに対応しきれないといった問題もある。特に、民間放送キー局の放送が優先されるため、地方局では受け皿がなくなる場合がある。

### ○実用化の状況（実用化の目途）

既に実用化されており、NHK、民間放送キー局・地方局におけるニュース番組、報道番組、バラエティ番組の他、特別番組、緊急放送番組等でも採用されている。

## 1.2.2 解説放送

### ①現状

解説放送は、NHK 総合で総放送時間の約 6%、教育で 10%程度実施されているが、民間放送局では大半の局で 1%以下という状況である。

解説放送が行われるジャンルとしては、ドラマ、バラエティが多い。

現在のところ、地上デジタル放送において、解説音声は全ての番組に付与できるわけではなく、設備上の理由により、解説音声を処理できない場合が 2 つある。1 つは、例えばステレオ 2 カ国語放送のように、2 つのステレオ音声を使用している場合である。この場合は、設備上、処理可能な音声サービス数は 2 つまでであるため、3 つ目の音声サービスとなる解説音声を付与することができない。もう 1 つは、5.1 チャンネルサラウンド放送番組の場合である。この場合は、設備上、別の音声サービスが処理できないため、解説音声を付与することができない。

解説放送の制作には、本編音声の隙間に解説音声を入れ込む台本作りをはじめ独特のノウハウが必要となる上、費用・体制面での負担が大きいことも課題となっている。また、解説放送の制作には 2 週間程度を要するため、放送直前に納品される収録番組には対応できず、結果として再放送の際に解説が付与される場合もある。ラジオドラマのスタッフやアナウンサーを起用して、スケジュールの空いたところで制作するケースもあり、その場合には放送用素材が 1 ヶ月前にできていなければ

難しい。

なお、視覚障害者のニーズが高い外国語のインタビューについては、主音声で可能な限り日本語の吹き替えを実施している局もある。

## ②技術開発例

### ○名称（開発主体）

解説放送番組制作支援技術（NHK）。

### ○技術開発の目的

視覚障害者向け解説放送番組制作の拡充・効率化。

### ○技術・システムの概要

平成 21 年度から、解説放送番組の制作者を支援する解説放送番組制作支援システムの研究を開始。高度な専門性が必要とされている解説放送番組を、経験豊富なスタッフでなくても効果的・効率的に制作できるよう、解説が必要なシーンについて画像認識技術を用いて自動検出する技術や、過去の解説台本をデータベース化して解説文の用例提示を行う技術の要求条件を決定した。

また、解説放送番組の台本作成に当たり、本編音声の有無を視覚化し、その隙間に解説音声を入力するといった作業を支援するエディターを開発した。

### ○技術・システムの特徴（従来技術との比較、効果）

音声を視覚化することにより、高度な専門性の必要となる解説放送番組の台本作成を効率的に行うことができる。

### ○課題

本システムが適用できる工程は解説放送番組制作の一部にとどまる。解説放送番組の制作には全体で 1 週間程度を要するが、そのうち本システムでは、6 時間を要する解説台本作成工程においてのみ、効率化を図ることができる。そのため、その他の解説音声収録等の工程においては、本システムが適用できず、全体的な効果が小さいことから、試験的に利用されたものの、採用には至らなかった。

### ○実用化の状況（実用化の目途）

現在、研究開発は行われていない。

### 1.2.3 手話放送

#### ①現状

手話放送は、現在 NHK 教育では総放送時間のうち、2%程度実施されているが、NHK総合や民間放送局ではほとんど実施されていないのが現状である。民間放送局でもわずかに手話放送を付与している場合があるが、費用や体制面で負担が大きく、現状を維持しているような状況である。このように手話放送の普及が停滞している最大の原因は、放送時に常に手話画面が表示されるため、画面が一部制限されてしまうことにある。

#### ②技術開発例

##### ○名称（開発主体）

CG による手話放送サービスに関する基礎的研究（NHK）。

##### ○技術開発の目的

聴覚障害者向け手話放送サービスの効率的作成・拡充。

##### ○技術・システムの概要

平成 21 年度より、日本語テキストからの手話への自動翻訳による手話 CG 放送サービスの技術的可能性について検討した上で、平成 23 年度よりアニメーションの自動生成の研究を開始した。手話を表現するための手指動作と顔表情のうち、手指動作について、手話通訳者の動きをモーションキャプチャーで録り、手話 CG モデルを構築した。また、日本語テキストからの翻訳に必要な日本語-手話対訳辞書を作成するとともに、類義語の使用により日本語語彙数を自動拡張する技術を用いて、約 4,900 語から約 86,600 語まで語彙数を拡張した。

図表IV-1-11 日本語-手話対訳辞書



資料：NHK放送技術研究所ホームページ  
(<http://www.nhk.or.jp/str1/open2010/tenji/flv/jpn/16.swf>) より

### ○技術・システムの特徴（従来技術との比較、効果）

アニメーションにすることで、画面サイズの拡大、アングル、キャラクターの変更といったアレンジが可能となった。また、TVML（TV program Making Language）や言語処理の研究の蓄積を活かして、例えば、手話通訳者の動作に、手指だけでなく唇の動きまで取り込む等、自然で高品位なCGモデルを構築した。

### ○課題

手話の文法等を表現するための表情については、ようやく研究を始めた段階である。

表示については、丸ワイプ<sup>4</sup>で画面の一部に表示する方法は評判がよくない。テレビ放送にインターネットで送信される情報を追加し、同時に画面に表示するハイブリッド・キャスト<sup>5</sup>等の応用を検討する必要がある。

### ○実用化の状況（実用化の目途）

今後5年程度で気象番組のような定型表現が多いものに対応できるアニメーション手話の技術基盤の開発が目標。

## 1.3 今後の字幕放送等の普及促進に当たっての課題及びその解決策

ここでは、今後の字幕放送等の普及促進に当たっての課題及び解決策を、主に生放送での字幕付与に関連して示す。

### 1.3.1 音声認識技術を活用したダイレクト方式・リスピーク方式における適用範囲拡大

字幕放送の普及促進のためには字幕付与のコストダウンが不可欠といえる。前述1.2.1で示した生放送での字幕付与方式のうち、リレー方式は相対的に多くの人員が必要であることから、それだけ人件費がかかる。そのため、音声認識技術を利用したダイレクト方式及びリスピーク方式の適用を増やしていくことが求められよう。ダイレクト方式及びリスピーク方式については、番組のジャンルによって発話者の特定／不特定、原稿や台本の有無、背景雑音の大きさ等が変わり、音声認識技術の特性上、認識精度に大きく影響するため、適用できる番組が限られてしまう。従って、適用で

<sup>4</sup> 画面上に丸型の小窓で映像を表示すること。

<sup>5</sup> デジタル放送システムに通信サービスを融合し、放送と通信が機能的に連携してテレビを軸にさまざまなサービスを展開するための基盤となるシステム。

きる番組の拡大が字幕放送等の普及促進に当たっての課題となる。

その解決策としては、前述 1.2.1(2)で示した通り、ダイレクト方式とリスピーク方式を組み合わせることでそれぞれの不得意分野をカバーした「ハイブリッド型字幕制作システム」の実用化を推進することで、適用範囲を拡大していくことが有効である。また、現在適用できていないジャンルの番組であっても、誤認識が一定程度少なくなれば校正作業の負担が軽減し、適用できるようになる可能性があるため、音声認識の精度そのものを向上することも重要となる。

また、本方式においては生放送中にリスピーカー、校正作業の人員が必要となるが、24 時間体制で人員を確保することは難しく、例えば早朝、深夜での番組や緊急放送が必要となった場合には対応が難しい。そこで、リスピークや校正作業の在宅対応を可能とするといったシステム開発等が求められる。

この点、海外では可能な限り字幕放送を実施するために、字幕付与作業の効率化を徹底している例がみられる。英国では多くの番組でリスピーク方式が採用されており<sup>6</sup>、最低限の人員で字幕を付与できる体制がとられている。英国では、字幕の正確さよりも、字幕付与番組を増やすことに重きが置かれているといえ、我が国においても、限られた予算のなかで効率的、効果的な字幕付与のあり方を検討するといったアプローチも必要であると思われる<sup>7</sup>。

### 1.3.2 リレ 方式における専門スタッフの育成支援

リレー方式については、前述 1.2.1(5)で示した通り、放送中に入力・校正等の運用人員が必須となるが、一定の専門的スキルを備えたスタッフの確保は容易でない。特に高速入力に対応できるオペレーターは限られており、また、その育成には数年を要することもあって、各局の需要に対して不足している状況にある。これに対しては、人材の需給バランスに考慮しつつ、専門スタッフ育成に係る検討が求められる。

### 1.3.3 地方局におけるインフラ整備支援

地方局においては、生放送での字幕放送等のための人員や設備を自ら確保できる局はごく一部であり、外部にも字幕制作の受け皿がほとんどない。従って、地方におけ

---

<sup>6</sup> BBC では 60%の番組でリスピーク方式によって字幕制作されている。

<sup>7</sup> ニュースなど特定のジャンルでは、報道機関がインターネットを通じて、テレビと同等かそれ以上の情報提供が行われている。テレビの字幕放送のあり方を検討する際には、こうしたインターネットなど新しいメディアの普及状況なども踏まえ、総合的に検討していくことも求められよう。

る字幕制作の人員と設備を整備することが課題となっている。このため、地域の放送局が共同で字幕制作を効率的に行う取組みの促進なども必要とされている。

**【参考：字幕制作会社を放送局が共同で設立した事例】**

字幕放送の普及による情報のバリアフリー化と障害者に対する新たな職域の提供による地域社会への貢献を目的として、福岡県の放送局5社と九州電力等が共同で平成16年7月に「株式会社九州字幕放送共同制作センター」を設立。九州地域の自主制作番組を中心に、30弱のレギュラー番組字幕付与を行っている。

#### 1.3.4 再放送時の解説放送の付与促進

解説放送の制作には2週間程度を要するため、初回放送のときに解説付与が間に合わなかった番組に対して、再放送の際に新たに解説が付与されることがしばしばある。これを踏まえ、解説放送の普及に向けて、まずは再放送時の解説放送の付与を促進することが考えられる。

#### 1.3.5 手話放送の表示／非表示を選択する機能の付与

手話放送においては、ワイプで手話画面が表示され、映像の一部が制限されることが普及促進の大きな課題となっている。そのため、字幕のように、視聴者側で手話画面の表示／非表示を選択できる機能が求められている。そのような機能を実現するため、前述のハイブリッド・キャストを可能とする規格づくりが求められる。



## 2. 通信における字幕付与

テレビ番組のインターネット配信に代表される、通信における字幕付与の状況について調査を行った。さらに、通信媒体での字幕付与の推進に当たっての課題を抽出し、運用面も含めてその解決策を整理した。

### 2.1 通信における字幕付与の状況

#### 2.1.1 通信における字幕付与の現状

NHK では、放送した番組を PC や高機能 TV 等に有料で配信する VOD（ビデオ・オン・デマンド）サービス<sup>8</sup>「NHK オンデマンド」を展開しており、PC 向けに計 15 番組（PC 向け全体の 13～15%程度）で字幕を付与している。

民間放送局でも同様の VOD サービスが行われているが、ビューアーに字幕データを重畳する機能が備わっていないことやニーズが把握できていないことから、配信番組での字幕付与が進んでいない。平成 23 年度に入り、字幕表示機能を備えた YouTube を使ったネット配信が一部で試行されている。

YouTube を使って配信する場合には 2 通りの方法がある。1 つは、局側から動画データと合わせて字幕データを提供し、その字幕データをもとに YouTube 側の字幕表示機能を用いて字幕表示する方法である。もう 1 つは、局側から動画データのみ提供され、動画の音声を YouTube 側で音声認識することにより字幕データを作成し、作成した字幕データをもとに字幕表示する方法である。

前者では、YouTube からの要請で字幕データの提供に協力するかたちとなっており、字幕データは生字幕あるいはネットやモバイル配信の際に作成したもの等を適宜利用している。他の用途に使った字幕データを利用することで字幕の正確性が担保できる点がメリットであるが、事前に字幕データを準備する手間がかかるというデメリットがある。一方、後者の方法では、事前に字幕データを準備する必要はないが、音声認識が不完全であることから、字幕のクオリティ・コントロールが難しいという問題がある。

#### 2.1.2 通信媒体で番組配信する際の字幕付与に関わる技術

NHK では、NHK オンデマンド（NOD）のリニューアルにあわせて、NOD 字幕自動変換技術を開発し、字幕付与機能を追加した。

---

<sup>8</sup> 利用者の要求に応じて、様々な映像コンテンツを配信するサービス。

### ○名称（開発主体）

NOD 字幕自動変換技術（NHK、エル・エス・アイ ジャパン）。

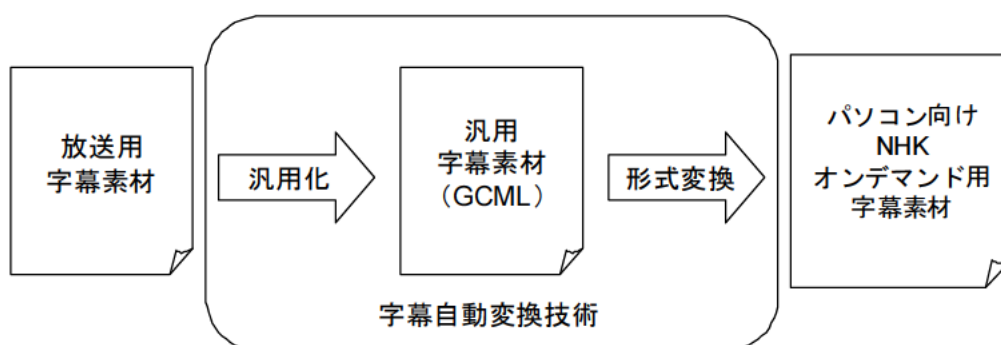
### ○技術開発の目的

放送で使用した字幕素材を NHK オンデマンドで利用できる形式に自動変換し、字幕付与の工程を効率化する。

### ○技術・システムの概要

放送用の字幕素材をパソコンの動画再生ソフトウェア（NHK オンデマンドではフラッシュビデオを使用）で表示可能な字幕形式 TTML（Timed Text Markup Language）に自動変換する技術を開発。テレビ放送と同様に、文字色、背景色、ふりがな、表示位置、点滅表示、下線表示、囲み文字等の設定ができるようにしている。

図表IV-1-12 字幕自動変換の流れ



資料：NHK報道資料（[http://www.nhk.or.jp/pr/marukaji/pdf\\_ver/264.pdf](http://www.nhk.or.jp/pr/marukaji/pdf_ver/264.pdf)）より作成

### ○技術・システムの特徴（従来技術との比較、効果）

変換技術で使用している汎用的な字幕言語 GCML（General Caption Markup Language）は、パソコンで扱いやすいXML（Extensible Markup Language）<sup>9</sup>構造であるため、字幕から抜き出したテキスト情報をメタデータとして活用できるようになる。これにより、番組のシーン検索等への活用の実現も期待できる。

### ○実用化の状況（実用化の目的）

平成 22 年 4 月より、1 日 4～5 本程度のドラマ番組においてパソコン向け字幕サービスの試行を開始。

<sup>9</sup> 目的に応じて文書やデータの意味・構造を記述するためのマークアップ言語の作成に当たって汎用的に使うことができる仕様。

### 2.1.3 課題

#### (1) クローズド・キャプション機能の付いたプラットフォームの構築

YouTube 等、通信を利用して配信する場合に、ユーザー側のビューアーで字幕の表示／非表示を選択するクローズド・キャプション機能が必要となる。字幕のあるものとないものの2通りのソースを提供することは局にとって負担が大きく、現実的ではない。民間放送各局においては、字幕放送や字幕データといった素材はあるので、クローズド・キャプション機能が備わった標準化されたプラットフォームがあれば、字幕放送の配信がしやすくなるとの意見が多い。

#### (2) 映像素材を加工した場合の字幕表示への対応

放送局が自ら VOD サービスにて番組を配信するに当たり、ネットでの配信が許諾されていない映像素材をカットしなければならないケースがある。また、有料サービスであることから一定の品質を担保する必要があり、番組中の間違っただけの内容等も同様にカットする必要が生じる。このように映像素材が一部カットされた場合、字幕を表示するタイミングも変わってしまう。そのため、改めてタイミングのズレを調整する必要が生じることから、現在のところ、加工の必要がなく、放送時と同じものを配信する番組でなければ、字幕付与は難しい状況である。

## 2.2 今後の字幕付与推進に当たっての課題及びその解決策

今後は、各局においては、過去に放送した番組を通信媒体で配信するケースが多くなり、それに従ってVODサービスにおける字幕付与へのニーズも高くなっていくものと考えられる。

放送局において、ドラマをはじめとして字幕を付与した番組については字幕データをもっているため、まずは、クローズド・キャプション機能をもつビューアーの普及を進めることが求められる。そうした環境づくりを進めるとともに、ユーザーとなる障害者、高齢者のインターネット利用の状況に鑑み、インターネット配信でのテレビ番組での字幕付与を、地上波放送と同様に位置づけることの検討も求められよう。

また、通信媒体で配信する番組に対して字幕付与が進まない原因の1つに、著作権等の問題により放送時に使用した字幕素材をそのまま利用できない場合には、インターネット配信向けに編集しなければならないことがある。このため、映像を一部カットした場合に、もとの字幕データから、映像カットに合わせた字幕データを自動生成するシステムの開発など、効率的に通信媒体向け字幕番組を制作する仕組みづくりが求められる。

### 3. 字幕データを活用した新たなビジネスモデル創出の可能性

テレビ番組に付与される字幕データの二次利用の状況について調査を行った。さらに、字幕データを活用した新たなビジネスモデルの可能性について検討した。

#### 3.1 字幕データの二次利用の状況

現状、字幕データの二次利用は積極的に行われていない。テレビ番組を DVD やモバイルコンテンツ、デジタルサイネージ向けコンテンツ（ニュース）として提供する際に、以前作成した字幕データを利用するといった一部での活用にとどまっている。

#### 3.2 字幕データを活用した新たなビジネスモデルの可能性

テレビ番組に字幕を付与することは大きなコストがかかることから、今後字幕付与を促進するためには、字幕データを活用することによって新たな収益機会が得られるといったことが求められる。こうした観点から、いくつか新たなビジネスモデルの可能性を検討した。

##### 3.2.1 モバイル端末/デジタルサイネージ向けコンテンツでの活用

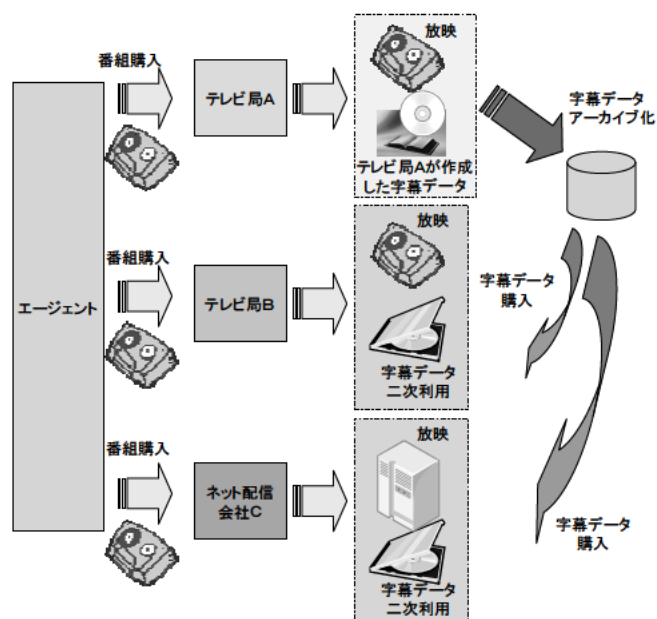
今後モバイル端末での動画視聴やデジタルサイネージの普及が拡大していくことが想定される。こうした用途向けにテレビ番組を提供していくに当たっては、屋外や公共の場でも音声を出力せず利用できるようにするために字幕は不可欠であり、テレビ番組の二次利用を積極的に進めているテレビ局では、モバイル端末やデジタルサイネージでの配信が想定されるテレビ番組には字幕を付与することが当然になっていくものと考えられる。そこで、モバイル端末やデジタルサイネージといった新しい媒体でのテレビ番組の流通を促進することは、テレビ番組への字幕付与が促進される側面がある。ただし、テレビ番組と、モバイル端末やデジタルサイネージで流通する番組では規格が異なるため、モバイル端末やデジタルサイネージ向けの番組にテレビ番組で付与した字幕データをそのまま使用できないといった問題があることから、規格の統一化を図ることが必要である。

##### 3.2.2 テレビ番組の二次利用時の活用

海外からテレビ番組の放映権を得た場合、日本語字幕が付与されていない番組には、放映時にテレビ局が字幕を付与しているが、後日、同じ番組を他のテレビ局が

放映する場合には、海外から、番組に字幕が付与されていない状態のものが納品されるため、再度、字幕を作成して付与することがある。このような重複した字幕制作を避けるため、字幕データをアーカイブ化し、その番組の放映時に再利用（字幕データ販売）することが考えられる。

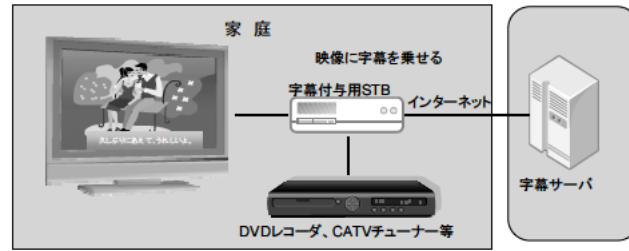
図表IV-1-13 字幕データのアーカイブ化イメージ



また、テレビの再放送だけでなく、インターネットを通じてアーカイブ化した字幕データを伝送することで、パソコン等ディスプレイで字幕付きの番組を利用することも考えられる。独立行政法人情報通信研究機構から助成を得てキューテック社は「聴覚障害者用字幕配信システム<sup>10</sup>」を開発、提供している。これは、DVDもしくはインターネット配信されているコンテンツに対して、字幕をサーバから配信、映像と同期させて表示させる技術である。また、テレビ画面にインターネット経由で取得した文字情報を重ね合わせて表示できるセットトップボックス (STB) がエンティス社によって開発されている。この STB を利用すれば、放送されているテレビ番組に、インターネット経由で取得した字幕を表示することができる。こうしたシステムの利用により、字幕付きの番組をいつでも利用することが可能となり、ユーザーの利便性向上につなげることができる。

<sup>10</sup> キューテック社が提供していた DVD に聴覚障害者用字幕を配信するサービスは、現在 NPO メディア・アクセス・サポートセンターに引き継がれている。

図表IV-1-14 字幕データのネット配信イメージ



### 3.2.3 スポンサーの獲得

字幕制作についてのスポンサーを獲得することで収益機会を得ることが考えられる。字幕番組を放映する際、番組の最初と最後に字幕にスポンサー名を表示することや、字幕の下部に常時スポンサー名を表示することなどによって、スポンサーの広告効果につなげることができる。

米国では、番組のスポンサーが字幕のスポンサーとなるケースと、字幕制作会社が番組スポンサー以外のスポンサーを募るケースがある。スポンサーを得た場合には、番組の最初と最後に字幕にスポンサー名が表示される。我が国でもこうした字幕に対してスポンサーをつけるといった取組は参考になろう。なお、民間放送各局で試行的に行われているCM字幕付与については、番組スポンサーにおける字幕放送への関心を高めるといった点から、テレビ番組への字幕付与を促進する上で有効であると思われる。

スポンサーの字幕表示については、スピードワープ研究所等があらかじめ用意したテキストデータを本編字幕の上部もしくは下部に表示する「副字幕」システムを開発しており、技術的には本編字幕の表示中に常時スポンサーを表示するといったことが可能となっている。

図表IV-1-15 副字幕によるスポンサー表示のイメージ



### 3.2.4 字幕データのメタデータとしての活用

字幕データをメタデータとして活用することが考えられる。テレビ番組において所望のシーンを呼び出す（検索）ことは非常に難しい。そこで、字幕データを利用して検索する。また、字幕データをPC等にダウンロードすることができれば、必要のところだけ見るといった活用の仕方もあると考えられる。