聴覚障がい　サポート

応用

ICT・音声認識の活用による講演・講義の字幕付与

5

河原達也（京都大学）

聴覚障がい者の情報保障の現状

2015 年春の本会全国大会を京都大学で開催するのに際して，新たな試みをいくつか行った．その 1 つが託児室の開設で，もう 1 つが聴覚障がい者への情報保障である．関連学会の状況を調査してもらったが，全国大会レベルではいずれもほとんど前例がなかった．聴覚障がい者の情報保障というと，手話通訳を用意すればよいと思っている人が意外に多い．しかし，手話（視覚言語）でコミュニケーションを行うのは生まれつき聴覚に障がいのあるろう者が中心で，はるかに多数の難聴者や中途失聴者は音声言語を中心にコミュニケーションを行い，手話を解さない人が多い．したがって，情報保障の手段は，音声言語を要約筆記・字幕化することになる．しかし，要約筆記を手配しようとすると，予想外に困難であることが分かった．京都のような規模の都市でこのような状況であることに愕然とした．

2011 年の東日本大震災・福島第一原発事故の直後の政府の会見やテレビの報道においても手話はつけられたが，字幕付与は人員の確保ができないという理由でほとんど行われなかった（その後，大規模災害時等緊急時放送について，できる限り字幕付与することが目標設定された）．

2013 年に制定され，2016 年度から施行される障害者差別解消法では，障がい者の社会的障壁の除去について「必要かつ合理的な配慮」を行うことが義務づけられている（民間の場合は努力義務）．「その実施に伴う負担が過重でない」範囲でという留保がついているが，学会や大学で対応するにははなはだ心もとない状況である．

今回の全国大会のもう 1 つの新たな試みが，メインのイベント企画のニコニコ動画による配信である．最近，MOOC をはじめとして，教育関係でも多くの動画コンテンツが作成・配信されるようになっている．現在テレビ番組の大半では字幕が付与されるようになったが，ネット配信のコンテンツでは字幕は皆無である．今回の全国大会の配信でも字幕を付加することを考えたが，これも容易でないことが分かり，断念せざるを得なかった．

このように情報保障のニーズと現状の人的資源や仕組みには大きなギャップがある．筆者らは，ICT，特に音声認識技術を発展させて，この状況を少しでも改善できればと考えている．

字幕付与の手段

音声をリアルタイムにテキスト化したり，字幕付与を行う手段を**表 -1** にまとめる．ここでは，速記者などのプロが行う特殊な場面と，一般の講演会などでボランティアが行う場面を分けて記載している．このうち手書きの場合は，多数の人への画面表示に向いていない．また原稿テキスト送出（前ロール）は，事前に原稿を用意して，読み上げることが分かっている場合のみ可能で，限られた場面でしか使えない．したがって現在，字幕付与に主に用いられているのは，タイプ入力である．テレビのニュースなどの生放送番組の場合はソクタイプを用い，校正者を含めて 4 ～ 6 名の大掛かりな構成になる．一般のパソコン要約筆記の場合も，2 名の連係入力で交代しながら行うので 4 名程度必要となる．要約筆記ボランティアの養成と確保が課題となっている．

字幕付与に，音声認識を活用する取り組みも行われている．ただし，発言者の自然な話し言葉音声を

	特殊な場面（プロの職業） テレビ番組・議会・法廷など	一般の場面（主にボランティア） 講演会や講義など
手書き	速記	手書きノートテイク
事前原稿利用	原稿テキスト送出	原稿テキスト送出（前ロール）
タイプ入力	ソクタイプ	パソコン要約筆記（PC テイク）
音声入力	復唱入力（リスピーク）	
直接音声認識	専用の音声認識システム	音声認識システムのカスタマイズ

表 -1　音声のリアルタイムテキスト化・字幕付与の手段

	実施状況	課題
パソコン要約筆記	第 1 イベント会場のみ	要員の確保が困難
手話通訳	要望に応じて手配	
手書きノートテイク	要望に応じて手配	情報量が少ない
直接音声認識	一部試行	精度の確保が困難

表 -2　全国大会における情報保障の取り組み[☆1]

すべて高い精度で自動認識するのは容易でないので，速記者やアナウンサが発話を復唱入力（リスピーク）して音声認識させる方式がある．復唱入力方式は，イタリア議会や米国の裁判所の一部の速記者が採用しているほか，NHK のスポーツ中継やバラエティ番組などでも使用されている．ただし，復唱入力には訓練が必要で，しかも誤りを修正する人やそれらの交代要員も必要となるので，一般の場面では試行例があるものの，あまり現実的でない．

これに対して，発言者の音声を直接認識する方式も研究開発が行われている．NHK 放送技術研究所ではアナウンサの音声を主な対象としてシステム開発が進められ，ニュース番組やスポーツ中継で実用化された．衆議院の会議録作成では，2011 年度から筆者らが開発した音声認識システムが採用されている．ただしこれらは，当該タスクの大規模な音声・テキストデータベースを用いて構築された専用のシステムで，直接一般の場面に利用できるわけでない．講演音声と書き起こしを多数収集した『日本語話し言葉コーパス（CSJ）』が構築されているが，個々の講演や講義では分野の専門性や講師の個性も大きいことから，十分な精度を確保するには，音声認識システムのカスタマイズ・適応が必要になる．

2015 年春の全国大会における情報保障の試み

以上の状況もふまえて，2015 年 3 月 17 ～ 19 日の本会第 77 回全国大会で準備した情報保障について**表 -2** にまとめる．パソコン要約筆記は，招待講演などを中心に第 1 イベント会場で実施した．イベント企画のみの聴講は当日参加も含めて無料であり，特に招待講演には一般の高齢者を含む多様な参加者が予想されるためである．ただし，要約筆記者を派遣してもらう団体では 1 日しか手配できなかったので，残りの 2 日は京都大学の障害学生支援ルームに依頼して，学生ボランティアで行うことにした．ただし，パソコン要約筆記ボランティアを多数養成できている大学は，全国でもまだ少ないと思われる（日本聴覚障害学生高等教育支援ネットワーク PEPNet-Japan[☆2] に登録している大学・機関は約 20）．

手話通訳と手書きノートテイクは，参加者から要望があれば対応することとした．手話通訳は 2 名から要望があり，京都市聴覚言語障害センターで手配した．手書きノートテイクは，障害学生支援ルームに依頼していたが，今回要望はなかった．

情報保障は，前回の全国大会からの申し送り事項

☆1 http://www.ipsj.or.jp/event/taikai/77/accessible_information.html
☆2 http://www.tsukuba-tech.ac.jp/ce/xoops/modules/tinyd0/index.php?id=27&tmid=4

にあったものである．実は，私の研究室には 1 年前に聴覚障がい学生がいて，ゼミではノートテイクをつけていたものの，彼が学会に参加する際に情報保障を要望することは思いもつかなかった．

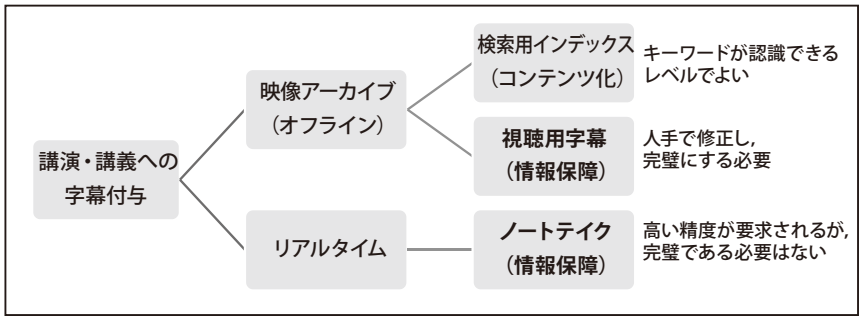

図 -1　講演・講義の字幕付与の形態

音声認識を用いた講演・講義コンテンツへの字幕付与

講演や講義の字幕付与には，**図 -1** に示すように，いくつかの形態があり，それによって求められる精度が異なる．

テレビ放送の事前収録の番組についてはほぼ全部に字幕が付与されるようになったが，インターネットで配信されている番組・コンテンツについてはほとんどされていない．表示するプレーヤや規格がさまざまあることの問題もあるが，一番の理由はコストであろう．YouTube では Google が音声認識による字幕を付与しているが，認識率は 50% 程度と報告されており，検索目的には利用できても，字幕として供するレベルではない．これに対して筆者らは，音声認識システムをカスタマイズ・適応することにより，高い精度の書き起こしを自動生成し，さらにこの結果を効率的に修正するエディタを含めた字幕付与システムを開発している．

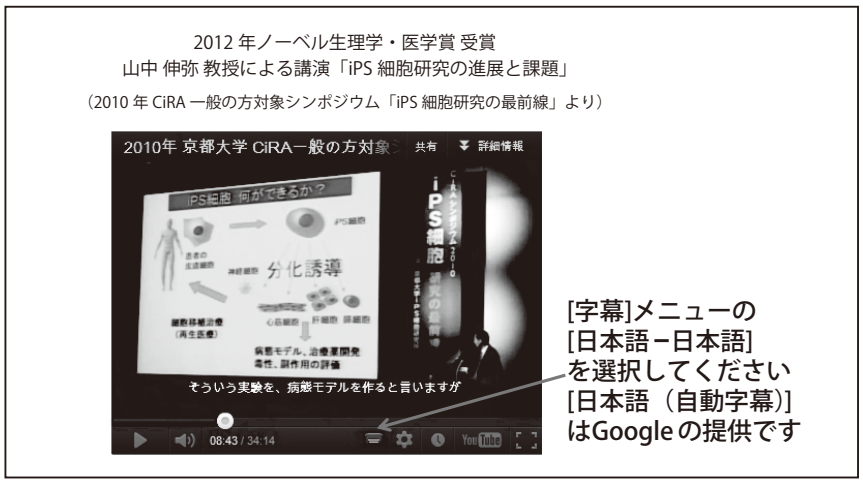

図 -2　京都大学 OCW の講演に対する字幕付与の例

京都大学 OCW[☆3]（OpenCourseWare）では数千件の講演動画が配信されているが，このうち，CiRA（iPS 細胞研究所）の一般の方対象シンポジウムと「大震災後を考える」シンポジウムシリーズの講演に対して取り組んだ．関連するテキストや講演スライドを用いて，音声認識の単語辞書や言語モデルを適応することで，60 ～ 85% 程度の認識率が得られた．認識率のばらつきが大きいが，最も影響するのは音響条件（マイクとの距離や雑音・残響）であり，収録時に留意してもらう必要がある．この音声認識結果を編集して作成した字幕を付与・配信している（**図 -2** 参照）．

放送大学で配信されている講義では，受講生の要望に応じて一部に字幕が付与されているが，全体の半分以下にとどまっている．そこで，インターネットで配信されている講義に対して，音声認識を用いた字幕付与に取り組んでいる．放送大学の講義は，科目ごとに 1 冊のテキストが作成されるので，音声認識の単語辞書や言語モデルの適応ができる．また，スタジオで収録されるので音響条件もよい．したがって，高い音声認識精度（80 ～ 95%）を実現できる．1 コマ 45 分の講義に対して，おおむね 4 ～ 5 時間で字幕の編集ができている．台本がある場合はさらに大幅に短縮できる．ここでの編集は，音声認識の誤りの修正に加えて，話し言葉の整形，句読点や改

☆3 http://ocw.kyoto-u.ac.jp/

行の挿入である．なお，音声認識の過程で，音声とテキストは単語単位で時間同期されている．

実際にコンテンツを視聴すると，聴覚に障がいがなくても，字幕によって理解が深まる印象がある．「病態」などの専門用語は漢字テキストで表示された方が分かりやすいし，視覚（字幕）と聴覚（音声）で入力される効果もあると考えられる．

音声認識を用いたリアルタイム字幕付与（ノートテイク）

日本学生支援機構[☆4]の調査によると，毎年約1,500名の聴覚障がい学生が大学等で学んでいる．各大学で講義やゼミでの情報保障・ノートテイクが行われている．その多くは手書きのノートテイクであるが，書く速度は話す速度より大幅に遅いので，「2割要約」と揶揄されている．一部では，パソコン要約筆記（PCテイク）も行われている．2名の連係入力により，ほぼすべての内容を字幕可能とされているが，設備や養成の問題がある．また大学の専門課程の講義や学会の講演では，専門用語が多いので，同一の専門の学生でないと作業が困難であり，要員の確保が容易でない．

音声認識システムは専門分野への適応が容易であるが，ネット配信されていないような一般の講演・講義では多様なスタイルがあり，認識率の確保が容易でない場合も多い．筆者らも何回か試みたが，実際の講義では音声認識率はおおむね60%程度であり，半分程度しか情報保障できていない．しかし，学会講演のような状況では，音響的条件がよければ実現可能と考えている．本会では今後，研究会や全国大会の講演のインターネット配信・アーカイブ化を行っていくので，字幕付与についても取り組んでいきたいと考えている．

いつでもどこでも字幕へ

これまで，字幕を作成・付与する過程について述べてきたが，ICTを用いることで幅広い配信の可能性が広がる．たとえば，パソコン要約筆記で一般に用いられているIPtalk[☆5]の出力を，Webサーバ経由でスマートフォンやゲーム端末に配信するシステムも作成されている．このシステムを発展させれば，インターネット動画配信と連携させることも可能になる．技術的・ポリシー的に解決すべき課題があり，今回の全国大会では間に合わなかったが，将来実現したいと考えている．

また，聴覚障がい者の日常生活におけるコミュニケーション支援のために，音声認識を用いたスマートフォンアプリの「こえとら」[☆6]やタブレットアプリのSpeechCanvas[☆7]，そして両方に対応したUDトーク[☆8]などが開発されている．これらもゼミや協働学習などの場で有用となる可能性がある．

本稿で紹介したテーマに関する最新の情報・意見交換を行うために，毎年京都大学で『聴覚障害者のための字幕付与技術』シンポジウム[☆9]を開催している．このシンポジウムでは，ソクタイプや音声認識を用いたリアルタイム字幕付与の実演も行っている．聴覚障がい者，要約筆記者，教育関係者，速記者，ICT研究者などが集まり，交流する場となっている．筆者は元来音声認識などを専門とする者であり，本稿で記載した内容の大半は本シンポジウムにおいて得たものである．参加・講演いただいた多数の方々に感謝する．

参考文献

1）嶺　重慎，広瀬浩二郎 編：知のバリアフリー[☆10]，京都大学学術出版会（2014）．
2）吉川あゆみ，太田晴康，白澤麻弓：大学ノートテイク入門，人間社（2001）．

（2015年1月30日受付）

☆4 http://www.jasso.go.jp/tokubetsu_shien/index.html
☆5 http://www.geocities.jp/shigeaki_kurita/
☆6 http://www2.nict.go.jp/univ-com/plan/applications/koetra/
☆7 http://speechcanvas.nict.go.jp/
☆8 http://udtalk.jp/
☆9 http://www.ar.media.kyoto-u.ac.jp/jimaku/
☆10 http://www.kyoto-up.or.jp/book.php?id=1992

河原達也（正会員）▌kawahara@i.kyoto-u.ac.jp

京都大学 学術情報メディアセンター／情報学研究科 教授．本会理事．音声言語処理，特に音声認識および対話システムに関する研究に従事．