

論文

聴者の主体的働きかけと技術の再配置による、 聴覚障害者—聴者コミュニケーション変容の可能性

——音声認識技術の利用検討から——

中 井 良 平*

1. はじめに

発話音声を変換する音声認識技術は、近年認識精度を向上させる技術革新を経たとされ、市場での存在感を高めている。その利用目的の一つが、機械的な情報保障と目されているのだが、情報保障場面での利用者側からの研究では、既存の聴者間コミュニケーションの枠組みに技術を配置するという考え方が中心となっており、聴覚障害者—聴者コミュニケーションについて問い直すという方向性にはなっていない。本稿では、坂本徳仁・櫻井悟史・立岩真也らの行った、生存学研究所における音声認識技術に関する過去の考察群を参照しながら、聴覚障害者—聴者コミュニケーションがどのように変容しうるかの考察を行う。手順としては、まず、先行研究に不足している領域を示すとともに、その原因について考察する。次に、最前線で高度にプログラムされた同技術と、無料アプリという対比的な利用場面を見ることで、同技術の限界と応用可能性について考察する。加えて、筆者が運営に参加したオンラインシンポジウムと、実際に聴覚障害者向けに提供されているアプリ利用場面について、それらが聴者のコミュニケーションの枠組みを出ない構造となっていることを確認し、改善策を提示する。

1-1. 先行研究

河原達也¹によれば音声認識技術研究の歴史は60年以上に及び、近年大きな技術革新を経た²。技術進展のあったこの10年に範囲を狭め、情報保障に関連する議論を見るために、聴覚障害者への言及があるものに絞り込むと、その先行研究はGoogle Scholarにてオープンアクセスできるもので300本程度となる。研究の傾向を見るために大まかに分類すると、以下ようになる。1) 技術についての研究。ここには、数値や計算式を用いて記述されたもの、技術の使い方、技術が利用された上でのフィードバックの記述等が含まれ、以上が先行研究全体の約6割を占める。ここでは情報保障やコミュニケーションのあり方についての議論は行われぬ。そして2) 教育の現場に関連する研究が、約2割を占める。ここでは実際の教育現場や学生／生徒のことが記述され、また想定され、現状の報告が行われ技術の利用可能性が語られる。情報保障のあり方についての考察が散見されるが、総じて、有限な予算や人手という条件内において、情報保障をどう工夫し行っていくかが苦慮されており、聴覚障害者—聴者コミュニケーション自体を見直す議論とはなっていない。次いで3) 放送分野での字幕生成に関するものが、1割程度ある。これは1)に含めてよいものの、数の多さから独立の項目とした。以上までで全体の約9割が占められることになる。残りの1割程度は、アンケート等調査報告書、医学論文、教育場面以外の現場報告、等々であり、それらと並び、4) デジタルテキスト情報の重要性を意識しながら現在の情報保障に対し独自の提言を行う、あるいは別の角度から見ようとする考察が見られるものの、数が少ないために、分類や系統立てた整理は困難である。以上の分類は筆者による簡易的なものであり、範囲の重複が起こり得るのだが、その場合、上記の典型例に近いと考えられる方へ分類した。

キーワード：音声認識技術、障害者のコミュニケーション、情報保障、障害の社会モデル、聴覚障害者

* 立命館大学大学院先端総合学術研究科 2020年度入学 公共領域

詳しく見ていきたい。2)の中には、二神麗子ら³のように、教育機関での音声認識アプリの利用例とその際の工夫や気付きが記述された論文があり、同様の場面で技術を用いる際の参考とはなるものの、文中で「高等教育機関において、聴覚障害学生から合理的配慮の求めがあった場合、障害の状態に応じて支援者側にとって過重な負担とならない範囲でこれらの支援を提供することとなる」⁴とある通り、やはり予算や人手の有限性という枠組みを超えた議論とはなっていない。そしてここではまた、現時点で製品の形で提供されている同技術について論じられるという意味で、技術自体ではなく技術の利用されるある形態について論じられており、その点においてもまた、限られた枠内にとどまる議論となっている。障害のある学生／生徒に情報保障を行わなければならない、同時に全ての学生／生徒に決められた枠組み内での教育を提供することも必須であるという教育機関にとって、「過重な負担とならない範囲」内でいかに情報保障を行うかという課題は大きなジレンマとなっていると考えられ、必然的に狭い範囲での議論とならざるを得ない、ということなのかもしれない。また1)、3)の分類からは、障害者－非障害者コミュニケーションについて主に論じたものは除いてある。以上のように、音声認識技術に関連する論文の大部分では、コミュニケーションのあり方に踏み込んだ情報保障への言及は、殆どなされていない。ゆえに、それらの研究では、音声主体のコミュニケーションの中に、同技術をどう配置するのか、という構造が前提されている。

1-2. 多様なコミュニケーションに向けて

4)に分類されたものを見ていきたい。高岡正と立岩は2010年に行われた「聴覚障害者の情報保障を考える」と題したシンポジウム⁵で、障害者のために提供される情報保障(傍点筆者)という、一般に非障害者が陥りがちな考えを批判し、非障害者の側が障害者に合わせたコミュニケーションをとることの論理的整合性について言及している⁶。また、大野口ベルト⁷は言語のユニバーサルデザイン化というアイデアで、情報保障と絡め、コミュニケーション方法の変容について考察を行っている。以上の議論はいずれも論文形式としては報告されておらず、コミュニケーション変容への提言というテーマが、まだ議論の深まっていない領域であることを伺わせるものとなっている。

高岡と立岩の発言が行われた前述のシンポジウムは、筆者が所属する生存学研究所の前身である、「グローバルCOE「生存学」創成拠点」の主宰で行われたものだった。シンポジウムでの報告の一部を紹介し、立岩の発言を引用したい。そこでは櫻井が、その前年の2009年に行われた、「障害学会大会2009」⁸における情報保障実験についての報告を行った。その実験とは、音声認識システム「アミボイス」⁹を用い、大会における講演音声の即時字幕化を行おうというものであり、その結果が芳しくなかったことを受け、立岩は以下のように発言している。

私が今一つ思っているのは、そしてこれは障害学の教えだと言ってもよいと思うんですが、その不出来なものに「健常者」の側が、少なくともその側も、付き合うというやり方もあるということです。まだソフトは未熟だ、機械は未熟だ、けれども、そういったものを使いながら、他方の健常者、健常者社会の側が工夫して、相手、ここでは聴覚障害者の側に合わせるというやり方もあり得ると思うんです。¹⁰

立岩は、『私的所有論』以来、人々の生存・生活に関わる制度や人々のあり方について多くの著作で言及しており、それらの論の基本的な立場は、現在不十分である制度や仕組みは変更されるべきであり、またかなりの程度変更が可能である、というものである。そのことを押さえた上で、立岩の上記の発言を見ると、歯切れの悪さのようなものを感じる。ここでは健常者の側がコミュニケーション方法を合わせるべきである、とはされず、そのような「やり方もあり得る」という控えめな提言となっている。この迷いのようなものはどこに由来しているのだろうか。次に、上記発言の後に行われたコミュニケーションについての言及を引用したい。

2人の間でコミュニケーションする時に不便なことがあって、そのために何か手立てを講ずる。それは、AさんのためでもあるしBさんのためでもある。でも、それをどこか理解を間違ってしまうと、Aさん、つまり具体的には聴覚障害者のためである。で、新たな負担がどうかこうとこかっていう話を私たちはしてしまうことがあるんだけど、それはちゃんと考えれば、基本から間違っている。¹¹

ここでは、障害者—非障害者コミュニケーションにおいて講じられる手立てについて、それが障害者のために提供されるものであるという考えが、明確に批判されている。それはコミュニケーションを行う人たちのためのものであり、一方のためのものではない、と言われる。立岩のこの指摘を敷衍すれば、様々に異なる者同士で構成される社会における多様なコミュニケーションは、社会の全員によって担保されるべきである、ということになるだろう。この指摘と、立岩の多くの論の背景にある、仕組みは変わるべきであり変えられるという立場を接続させることは、難しくないように思われる。では、先の立岩の「健常者が障害者にコミュニケーションのやり方を合わせることもあり得る」という問題提起は、我々の認識が間違っており変更されるべきである、という論に対しどのような位置にあるのだろうか。その手がかりは、当時の音声認識技術が現在よりも極めて不完全なものであったという点にある。坂本らの報告によれば、当時の技術は導入に数百万円を要しながら、その水準は現在と比べ極めて低いものであり¹²、個別のコミュニケーション場面での現実的な利用可能性は皆無であったと言っていい。よって、立岩がそのような場面でのコミュニケーションの変容の可能性について言及した時、技術の利用可能性の延長線上にコミュニケーションの変容可能性を見ていたというよりも、コミュニケーションの変容可能性についての思考が技術を捉えたと見る方が妥当だろう。つまり、個別の技術はあくまで副次的なもので見られている。とすると、立岩の障害者—非障害者コミュニケーションへの言及の際想定されていたのは、特定の技術から導かれるコミュニケーションという経路ではなく、より多様であるべきコミュニケーションという考えから発した未特定の経路であり、そのことが言及に見られた歯切れの悪さのようなものに反映されていたと考えられないだろうか。そう考えた時、あるべき多様なコミュニケーションという問題提起に続いて現れる、ではどのように変容させれば良いのか、という思考領域の輪郭が見えてくる。

1-3. 技術提供者／利用者間の隔たり

立岩の障害者—非障害者コミュニケーションへの言及を念頭におき、改めて音声認識技術に関する先行研究を振り返りたい。技術について記述することに重点の置かれた研究が、障害者の参加可能なコミュニケーションについての考察を欠いていることは、その記述の形式上やむを得ない部分があるにしても、情報保障について記述した教育現場に関わる多くの研究も、同技術を用いたそのようなコミュニケーションのあり方を十分に論じられていないのは、教育現場における情報保障が有限な枠組み内にとどまらざるを得ないと仮定したとしても、奇妙である。障害者への情報保障を巡っては多く議論の蓄積があるはずであり、立岩のように、コミュニケーションから疎外された障害者への情報保障の手段として技術を見るだけではなく、非障害者側が、障害者も参加可能なコミュニケーションを構築するためにその時点で使える技術を使う、という方向性へ議論が進んでもいいはずなのだが、現状はそうにはなっていない。そして、そのような議論は、教育に関係した研究として行われる必要はないはずなのだが、現状はそれ以前に、音声認識技術を利用する側が記述した研究は、教育関連のものに偏っている。その理由として、考えつくことのできる可能性の一つは、同技術を開発・運用する側と利用する側の情報の断絶、交流のなさである。両者の間に交流が少ないために、技術の提供側には障害者や障害者への情報保障を行う側の抱く問題意識が、技術を必要とする側には技術に関して十分には伝わっていないのではないか、ということである。前述のシンポジウムで櫻井らは、その前年2009年の9月に行われた障害学会大会において、音声認識システムを用いて行われた情報保障実験についての報告を行っているのだが¹³、その試みに関連して、技術提供側と利用側の交流の不足を示すことを示唆すると思われる記述を、河原・秋田祐哉¹⁴に見ることができる。河原は2009年11月に行われた「『聴覚障害者のための字幕付与技術』シンポジウム2009」における音声認識技術を用いた字幕生成が、「一般の講演会で直接音声認識による情報保障を行った国内最初の事例」と考えられるとしている¹⁵のだが、前述の坂本らの実験でも同様の試みが行われており、それは河原らに2ヶ月先駆けてのものであった。坂本らが情報保障を専門に研究する者ではないために、河原がその報告を見落としした可能性も考えられるのだが、坂本らの研究が河原に届かなかったこと自体が、両者の交流の不足を示しているとも考えられる。その仮説をさらに進め、隔たりの原因について考えた時、河原をはじめ情報保障に関心を寄せ継続的に研究を行っている者達が少なくないことを勘案すれば、技術を利用する側にいる研究者達が、技術自体にそれほど関心を抱いていないのではないか、という考察を行い得る。音声認識技術を用いた日本初の情報保障の試みという先駆的な取り組みに関わり、障害者—非障害者コミュニケーション

の変容について言及を行っていた立岩やその門下達にしても、その後同技術に関連した研究を行っていない。そのことは、重要であるのは人々のあり方であり、技術は人々を仲立ちするピースに過ぎないという立岩らの研究姿勢を表していると考えられるが、それはまた同様に多くの社会科学系の研究者の研究姿勢と言えるのではないだろうか。技術の利用者側の研究に偏りがあることも、技術を使わざるを得ない場面に居合わせた研究者達が、主体的というよりも必要に迫られてその記述をはじめたと考えれば、納得がいく。

ここまででは、音声認識技術を用いた情報保障に関連した先行研究において不足していた領域についての指摘と、その原因についての考察を行った。そこには、かつて立岩が言及しそのままになっている、音声認識技術を媒体としたコミュニケーション変容の可能性、という考察が含まれていたことを確認した。立岩の問題提起を引き継ぐ場合、本来、その媒体は広く技術全般となるはずである。しかしながら、既に見たように、音声認識技術に関しては利用者側による利用検討が深まっておらず、生存学研究所の過去の考察群を引き継いだ上で、現在の技術水準における可能性と課題を検討する必要があると考え、本稿は書かれる。その際、製品として提供されている技術の枠内ではなく、現時点で容易に応用可能な技術を想定して考察を行う。それは技術の専門家でなくとも行える考察でありながら、先行研究では殆ど行われていない作業である。技術に関連してそのような考察を行える余地が広く存在するということが、技術と技術を利用する・必要とする人々に遠い距離があることを示していると考えられる。

2. 音声認識システムの利用実態

市場で広く提供されている音声認識技術であるが、技術提供元による案内や利用者のレビューから、技術が有用なものとなりはじめていることは分かるものの、最新の技術にどこまでが可能であり、何が課題なのかといった、専門性に踏み込んだ疑問については、素人向けの大雑把な回答、または、専門家向けの部分詳細な研究ばかりである。可能性と課題について正確に知ることは、技術自体の利用検討を行おうとする際、必須の作業である。ここでは、同技術が高度にプログラムされ利用されている最前線としての、各国のテレビ放送における字幕付与場面を見ることにより、現在の同技術の限界を明らかにする。また、エンドユーザー向けの無料アプリ、「UD トーク」を用いて簡単な実験を行うことで、同技術の応用可能性を探る。

2-1-1. 日本国内の利用事例 (NHK)

佐藤庄衛¹⁶によれば、NHK 総合における 2016 年度の字幕放送の割合は、97.4%となっており、異なる 5 つの字幕付与方法が採用されている。そのうち、音声認識技術による自動字幕生成と呼べるものは「ハイブリッド方式」と呼ばれる方式中の一部分である¹⁷。それを可能としているのは、最新記事のデータベースとの連携による固有名詞や人名の自動学習であり¹⁸、独自の技術開発を行った NHK ならではのシステムと言えるだろう。それぞれの方式は、適した利用場面ごとに振り分けられ用いられており、事前収録番組に関しては人の手により字幕が付与されているとのことである¹⁹。事前収録番組は、生放送番組と比べ、字幕付与作業時間に制限がないため、より確実性の高い人の手による入力を選択されているものと考えられる。

2-1-2. 米国における字幕制作業者の字幕付与方法

井部千夫美²⁰によると、米国でも日本と同様、事前収録番組と生放送番組で、字幕付与の方法が異なるとのことである。字幕制作業者は、事前収録番組においては人の手により字幕を生成し、生放送番組においては、Stenotype と呼ばれる高速入力キーボード及び音声認識システムを用いていると報告されている²¹。

2-1-3. チェコにおける公共放送での字幕付与

海外におけるテレビ放送の情報保障について、詳細を知ることのできる機会は多くないが、チェコ公共放送において音声認識システムによる字幕付与が行われるようになった経緯と、その様子が記された論文として Pražák らのもの²²がある。いわく、主要な言語に限って開発されている高速入力キーボード (Stenotype, Velotype) はチェコ語の速記には利用できず、音声認識システムがリアルタイム字幕生成における唯一の手段だった。チェコには、

同様の状況に置かれた他国と異なり、音声認識技術の研究を行ってきた複数の大学があったため、同技術をリアルタイム字幕生成に用いる事が可能となった。リアルタイム字幕生成は、複雑な訓練を受けた復唱者によって行われ、復唱者を置かず、直接入力で字幕が生成されるのは、チェコ共和国議会のミーティング放送のみで、議会用に特別に調整され続けているそのシステムの認識誤り率は6%未満ということである²³。

2-1-4. 最前線の利用場面から見る技術の限界

以上までに見た、各国テレビ局での音声認識技術利用例から、同技術の利用について、次のように結論づけられると考える。1) 長く複雑な会話を文字化する場合は人の手による作業が最善手である、2) 生放送の場面など、字幕生成時間が限られている場合は、音声認識システムが用いられる場合がある 3) 設備投資による高度なプログラムを用い、訓練を受けた復唱者を置いても、会話の全てを機械的に文字に変換することは困難である。

テレビ局において独自研究された音声認識システム及び関連技術のレベルは技術の最前線と考えられ、上記が現在の技術の限界の重要な例と考えられる。

2-2. 技術の末端利用場面での実験

次に、技術の性能についてメーカー側がどのような説明を行っているのかを確認しておきたい。前述の坂本らの実験で用いられたアミボイスを開発する、アドバンスト・メディア社の、音声認識技術についての質問・回答集²⁴を参照すると、認識率を高めるために必要な主な要件は、概ね a) 雑音のない環境、b) 話者とマイクの適正な距離、c) アナウンサーのように明瞭に話すなど、話者が認識されやすい発話を行うこと、の三つとなる。曖昧な部分の多い条件となっているが、「アナウンサーのように明瞭に話す」という発話技術に関する条件は、実現のハードルが高いように思われる。そこには例えば、方言を用いず教科書通りに話すことも含まれている。では、認識精度を高めるためにはアナウンサーのような発話技術が必須なのかといえ、実はそうともいえない場合がある。メーカーは、普通の会話速度で話すことが認識精度を高めるとアナウンスしているのだが²⁵、鳥羽祐輔ら²⁶では、ある程度ゆっくりと丁寧話すことで認識率が上がると報告されている²⁷。これは、アナウンサーのように通常速度の発話に明瞭性を持たせることができずとも、明瞭さを維持できる速度で発話する、という方向性と考えられる。以下に、実際の認識例を示す。用いた文章は、本年の障害学会大会の運営にあたり、運営側から障害学会理事会に向けてメール上でアナウンスされた大会説明²⁸から、話し言葉に近いものを抜粋した。認識には、音声認識システムを備えたスマートフォン用アプリである「UD トーク」を用いた。

原文：

私にとってはちょっとうちわかんありすぎなんです、それは私自身がある程度のことを知っているということでもあるので、聞き手になって話をさせていただくにはよいのかなと思います。

筆者発話の認識結果（通常速度で発話）：

私にとってはちょっとうちはかなり席なんです、それは私自身がある程度のことを知っているということでもあるので、規定になって話をさせていただくには良いのかなと思います。

筆者発話の認識結果（ゆっくり丁寧に発話）：

私にとってはちょっとうちは完ありすぎなんです、それは私自身がある程度のことを知っているということでもあるので、聞き手になって話をさせていただくには良いのかなと思います。

通常速度の発話では、話し言葉的な色の強い、「うちわかん（内輪感）」及び、比較的スムーズな発話の難しい「聞き手」という単語が認識されなかった。それらは丁寧に発話することにより、「うちは（わ）かん」と正確な音で認識され、「聞き手」は音・漢字とも正確に認識された。文節を区切るよう意識し発話することにより、読点も正しい位置に振られていることがわかる。システムは登録済みの単語辞書を参照して認識を行っており、登録のない単語

は認識しないため、「内輪感」が誤変換となったと考えられる。また、システムは前後の文脈から解析・認識を行っており、助詞を正確に用いることも重要である。つまりは、話し言葉や独特の言い回しを避け、教科書通りに丁寧に話すことで、認識率が上がるようになっている。通常の話し言葉では、以上のような点に十分に注意を払った発話が困難となり、そのことも、明瞭さとは別の要因として、認識率の低下につながっていると考えられる。

以上のように、適切な発話速度で、明瞭かつ平易に話せば、コミュニケーションの場における音声認識システムの活用可能性は、決して低くはないと考えられる

3. 音声認識技術の利用検討から考えるコミュニケーション変容の可能性

以上までに見た通り、音声認識技術の利用にあたっては注意点が多くあり、高度にプログラムされた同技術とシステムを連携させ利用している各国のテレビ局においても、人による従来の情報保障を効率化あるいは、部分的に置き換える程度の利用にとどまっているということがわかった。市場で提供されている水準の技術を用いた場合には、必然的に、それらの場合よりも質の劣る情報しか生成されず、人による大幅な編集作業が加わらなければ、人の手のみによる情報保障の代替とはなり得ないということになる。以上は、話者が速度等を気にせず発話した場合であるが、一方で、認識が行われるよう、話者の側が積極的にシステムに合わせて発話すれば、認識制度がかなりの程度高められることもわかった。

3-1. 音声認識技術を用いた聴覚障害者—聴者コミュニケーションの再提案

ここで1節の高岡と立岩の問題提起に立ち帰りたい。音声認識技術による字幕等の生成は、聴者のためという部分もあるものの、基本的には聴覚障害者への情報保障という側面に重点がおかれ考えられている。そこでは、聴者への情報提示のあり方、つまり音声の利用が可能かつ有効である場合、音声の主體的に用いられ、文字情報は利用されないか付随的なものにとどまる、という状態（テレビ放送、公演、会話の場面など）が前提とされており、そこに情報保障が加わることになる。すなわち、聴者同士の音声コミュニケーションに、聴覚障害者が受け取れる情報を追加で生成するという考え方であり、その場合、情報保障が行われたとしても、聴者同士のコミュニケーションはそのまま据え置かれる。音声認識技術に関する先行研究の殆ども、このような構造を前提として書かれており、情報保障という考え方自体が、そのような特性を持っていると言えるだろう。他方、立岩が重きを置いていただろうと1節で考察した、その変容という方向性で聴覚障害者—聴者コミュニケーションを考えた時、音声認識技術の利用可能性について、より広く考察が可能となる。

先のシンポジウムで立岩は、必要なすべての場面に通訳者が配備されるべきであるという、情報保障的な方向性を基本的な路線としながら、次のように言及していた。

例えば私が少し、口の動かし方とかを練習して、そして、タイムラグが生じるのであれば、私自身がその画面を見ながら、「あ、だいたい表示し終わったな」と思ったら、次のセンテンスに進むっていうやり方もありだと思うんです。機械を使いながら、喋っている側が何か工夫をして、スピードを落としたりしていけば、私とその機械の二人三脚で少なくともスクリーンやディスプレイに映していく、そういう方法というのも一つにあるのではないかなと思うんです。²⁹

これは一見、あるシステムを利用する際の些細な工夫のようにも感じられるが、情報発信者である非障害者の側がコミュニケーションに必要な情報生成を自ら（ここでは機械の力を借り）生成するという点で、情報発信者に限らない者が既存のコミュニケーションに情報を付随させるという、情報保障というテーマに中心的な考えとは明確に異なるものになっており、それゆえに、情報保障に関する既存の研究では見落とされがちな視点であったと考えられる。しかし、当時の音声認識技術は、立岩の提案したような用い方ができる水準にはなかった。既に見たように、認識精度がかなり低く、システムの導入には数百万円という高額なコストがかかった。1節で述べた通り、その後立岩らは音声認識技術に関する研究を停止してしまったのだが、障害者のコミュニケーションに関する問題意識が立

ち消えなくなってしまったわけではなく、例えば立岩が会長を務め、筆者が運営作業に参加した、障害学会の2020年度大会³⁰における報告では、音声認識技術の利用が検討された上で、同技術を使わずにコミュニケーション方法を変更するという対応がとられた。櫻井らの研究と立岩の言及の後の10年で、音声認識技術は認識精度を飛躍的に向上させ、コストも安価になったため、そこでは当時は行い得なかった利用検討が可能となった。次項以降で、当時の問題意識を受け継ぎながら、音声認識技術を媒介としたコミュニケーション変容可能性について、再提案を行いたい。

3-2-1. 障害学会大会2020での利用検討

障害学会大会は、障害学会が毎年秋頃に行っており、2020年度は、COVID-19の影響ではじめてのオンライン開催となった。筆者は今年度の運営作業に関わり、同大会におけるオンライン報告場面での音声認識技術の利用検討などの作業を経験することになったため、そこで得た知見と考察を以下に記す。例年、同大会では「ポスター報告」という、ポスターによる報告をボードに張り出し、報告者がその前に立ち質疑応答を行うという形式の報告を、プログラムの一部として行っていたのだが、2020年度については、オンラインに報告全文を前もって掲載しメールで質疑応答を行うという形式へ変更が行われた。そのことにより、従来会場において音声と視覚情報を用いて行われていた報告が、オンライン上で期間中は制限時間なく、全てデジタルテキスト情報を用いて行われることとなった。ここでは、音声のみ、視覚のみ、という単一の受け取り方しかできない情報の提示方法が排され、ディスプレイへの表示、読み上げソフトでの対応、デジタル点字への変換等、複数の方法での情報受容が可能となる、デジタルテキストという媒体が選択されたということになる。その決定の経過は、以下のようなものだった。

当初はウェブ会議ソフトで各報告者が報告し、そこに人による通訳もしくは音声認識システムを用いた情報保障を付加するという方向性であったのだが、予算の都合で通訳がつけられないということ、また、音声認識システムを使おうにも訓練を受けた人を運営側で用意できない、ということになった。そこで立岩の発案により行われた決定は、オンライン会場での音声報告を取りやめとし、前述のデジタルテキスト情報のみによる報告、及び、メールを用いたデジタルテキスト情報のみによる質疑応答を行う、というものだった。そこで行われたのは、通訳などを用意し、非障害者のコミュニケーションに障害者が参加できる別の情報経路を設けるという方向から、特別の情報提示がなくとも、障害者（そこで想定されたのは、視覚・聴覚障害のある人だった）でも参加できる形態のコミュニケーションとする、という方向性への転換であった。音声認識技術が用いられなかったことと、そのような方法がとられた理由についてももう少し詳しく述べたい。同技術を利用する場合は、運営事務局内部で人員を配置する構想であったのだが、事務局内に技術利用に慣れた者がおらず、専門家である通訳へ依頼する場合に近い情報保障は行えないだろうと考えられたことが、同技術を使わないと判断された大きな理由であった。その上で前述の方法となった理由は、報告に関しては、報告者による学術的な情報提示がその大部分になると考えられ、掛け合い等、即時発話の必要度の低いコミュニケーションであると判断されたためである。同報告形式について、大会の終了後筆者は、会場で行われた以前の形式の大会でポスター報告を行い、今回の大会にも参加された、視覚障害者であるAさんから、報告形式についての感想を聞いた。Aさんからは「従来の方法では他の報告を回る時間がなく、メインのシンポジウムしか聞くことができなかった」が「今回はゆっくりと他の報告を読むことができ、かえって良かった」と評価頂いた。一方、大会のメイン部分であったシンポジウムに関しては、ウェブ会議ソフトを用いて、専門家による手話・文字通訳のリアルタイムでの配信が行われたのだが、手話通訳者画面の表示サイズが小さいことをはじめ、次回以降の改善を必要とする複数の意見が寄せられた。原因として、運営側に障害当事者が入っていなかったこと、並びに、学会大会のオンライン開催が運営側にとって初めての作業のため、不慣れな部分が多かったことが考えられる。

3-2-2. オンラインシンポジウムにおける音声認識技術の利用可能性

今回のオンライン大会で音声認識技術を利用できる可能性があったか、また、その際の条件や課題はどのようなものであるか、という点について若干の考察を行いたい。この10年の技術の変化は既に述べた。とすると、立岩が提案したような、話者の工夫によりコミュニケーションに技術を組み込むといったことが可能になっているのだろうか。ここでは、同技術の利用は、専門家の手を入れず、原則話者本人が行うという方向性で検討する。その場合、

認識精度を高めるための各自の工夫が必要となり、発話速度は聴者同士の場合と比べ遅くなると考えられる。

まず、それが音声によってであれ、他の手段によって行われるものであれ、当該のコミュニケーションがどの程度即時の反応を必要とするものであるのか、という点が考慮されるべきだと思われる。例えば漫才のような掛け合いは手話でも行われているが、ゆっくりと情報を生成しそれを待ってから次の者が同様の作業を行い、それが繰り返される、という方法には向かないようにも思われる。翻って、シンポジウムではまず各話者が置かれており、それぞれある程度の時間を使い話すことになるため、その部分に関しては同技術を利用し、各話者に情報生成を行ってもらおうという対応が可能ないように思われる。細かなニュアンスや専門用語については第三者による即座の修正判断が難しい場合があるので、音声認識ツールと同期可能な字幕生成ツールを用い、話者自身に適宜修正してもらおうという対応が考えられる。質疑応答に関しては、メール等を使って受け付け、それを回答者が読み上げるという方法で、話者の入れ替わりを少なくできそうである。話者が固定された長時間の講話を想定した場合、同技術の利用可能性は比較的高いように考えられ、そしてオンラインでのシンポジウムというのはそのような形態であるだろう。課題はむしろ技術よりも、それを用いる人間の側の足並みにあるかもしれない。ここで目指されるのは、ある程度の情報保障ではなく、各自が責任を持ち情報を発信してコミュニケーションを行うことである。各話者は技術をまた、自身のコミュニケーション技術の一部としなければならない。その際、音声スムーズに認識される人もいれば、何らかの要因によりそうならない人もいよう。例えば、方言の利用者は教科書通りの話し方に変える必要があるのだが、それが実際に可能であるのか、という問題が想定される。通常の講演では行われず、本番形式での繰り返し打ち合わせや、各話者が長時間の発話に対し実際に技術を用いて文字情報を生成するテスト及び、一定の認識率となるまで試行錯誤を繰り返すといった時間も必要になってくるだろう。技術が使えることはある程度わかっている、しかし利用にあたっては人間側が機械に合わせる必要がどうしても出てくる、場合によってはかなりの努力が必要とされる人がいるかもしれない、ということである。あるシンポジウムという場合を想定しても、一ター朝にはことは進みそうにない。そして、ここでそのようなコミュニケーションの変容を迫られているのは、テキスト情報をさほど必要としない聴者の側である。

以上のように、この10年の技術進展ののち、オンラインシンポジウムの場面で音声認識技術を用い、立岩の提案を実現しようとした際に想定される課題について、考察を行った。

3-3-1. 「みえる電話」 利用場面における課題

次に、聴覚障害者の利用を想定され、スマートフォン利用者に提供されている、音声認識技術を用いたサービスである「みえる電話」におけるコミュニケーション場面を、障害者—非障害者コミュニケーションの一場面と捉えた上で、聴覚障害者の観点から見た場合の問題点と、その技術的な対応に関する考察を行う。

「みえる電話」サービスは、NTTドコモと契約したスマートフォンで利用できるサービスで、聴覚障害者として想定されている利用者がスマートフォンを用いて、通話相手の発話音声は音声認識技術でデジタルテキスト化された文面を読み、テキストを手動で入力し返答したものが相手方に合成音声で再生される、という仕組みになっている。

音声認識技術を用いた同サービスについて、提供元であるドコモの公式サイトを参照しながら、詳しく見ていきたい。なおここでは、騒音の多い環境で同サービスを利用した場合を想定し、考察を行う。既に見たように音声認識技術は、限られた条件下でしか高精度の音声認識を行えず、騒音の多い環境等では、役に立たない場合があるというものだった³¹。

同サイトでは、利用を想定した各場面がイラストで紹介されている。レストランの予約、病院への問い合わせ、緊急時の連絡、宅配便に関する問い合わせ場面のイラストがあり、そこではいずれも、スマートフォンを操作する聞こえない利用者と、固定電話から発話する聴者が描かれている。イラストの内容から分かるように同サービスは、聞こえない利用者がスマートフォンとアプリを用意し、聴者は特段の準備を必要としない仕組みになっている。そこから、騒音の多い環境等での利用時について推測できることは 1) 聞こえない利用者が手動入力で相手に呼びかけることは可能だが、固定電話側は応答手段を持たない、2) 聞こえない利用者側から音声認識が上手くいっていないことを伝えることは可能だが、そのことが伝えられなければ、固定電話側は、音声認識の成功／失敗を知ることができない、3) 音声認識が上手くいっていないことを固定電話側が知り得たとしても、認識結果を参照できるわけ

ではないので、大まかな対応しかできない、4) スマートフォン側と異なり移動ができないため、固定電話側は騒音の少ない環境へ移ることはできない、等である。以上のように、同サービスは、コミュニケーションの妨げとなるいくつかの大きな問題を抱えているものであることが分かる。

これらの問題を解決するためには、発話者へ音声認識結果のフィードバックが行われること、及び、代替コミュニケーション手段への移行がスムーズに行われるようなシステム設計がなされることが必要となってくるだろう。具体的な方法についての考察は割愛するが、当該コミュニケーションで既に用いられている合成音声技術によるアナウンス機能を用い、また、聴者側の多くが利用しているスマートフォンとの連携を行えば、同様のシステム実現に際しての大きな技術的困難はないはずである。通話の代替手段については、双方によるテキスト入力が想定されるだろう。以上のような方式を用いても、代替コミュニケーションが行われ得る範囲は限られたものになるであろうが、それにしても、スマートフォンの普及率を考えれば、それはかなりの広範囲となるはずである。テキスト入力に関しては、タッチパネルの操作ができればよく、ATM 等で同等の操作が広く行われていることを考えれば、適切な操作指示が行われるよう工夫がなされれば、使い慣れていない人にとっても、そう大きなハードルと考える必要はないかもしれない。

もしそのような対応が迂遠に感じられるとすれば、障害の社会モデルに則り以下のような反駁が可能である。すなわち、当該ケースにおいては、初期状態では聞こえない利用者が、聴者の通信手段である電話の利用を強いられているのであり、テキストでのやり取りに移行したとしても、両者が同条件でやり取りを行うことになったに過ぎない。そのように、現状の不均衡を均し、不足部分を補う方向性で考えれば、「みえる電話」のようなサービスが、より双方向性を持ったコミュニケーションとなる可能性を考えつくことができる。

また、この案に対し、聴者の側がスマートフォンを用意することや、テキスト通信を可能にするアプリ導入の困難が、反論として言われるかもしれない。しかしそのことは、聞こえない利用者側が現在それらの負担を負っていることを証明するに他ならない。例えば、全てのスマートフォンに必要なアプリをプリインストールするという方法をとれば、アプリの導入についての問題は解決するだろう。

3-3-2. 電話機を用いたコミュニケーションの可能性

次に「みえる電話」のようなサービスを発展させた場合、どのようなコミュニケーションとなり得るのかについて、考察を行いたい。現時点では同サービスは、聴者の側は負担を負わず、情報発信に際してもフィードバックがないため、行う工夫は限られるという仕組みになっており、障害者—非障害者コミュニケーションにおいて、非障害者の側のコミュニケーションは据え置かれるという、本稿で問題にしてきた構造と同一のものになっている。また、聴覚障害者側の負担に関しては、以下のような指摘も行い得るだろう。すなわち、そもそも、通話という手段が聴覚障害者を排除したコミュニケーションであり、音声認識技術を用いても不親切なものに過ぎず、聴者の側はそのことを意識し、より根本的な配慮を行うべきで、といった指摘である。その方向性で一つ考えるならば、聴覚障害者—聴者の電話機等を用いたコミュニケーションは、テキストによるやりとりを基本とするべきであるという案をすぐに思いつくことができる。確かにそうなれば、筆者が前述した、通話機能を用いたのちテキストのやりとりに移行する、というプロセスが省かれるように思われる。初めからテキスト同士のやりとりが行われるのであれば、そこには音声情報は介在せず、コミュニケーション手段は双方に同じく利用できるものとなり、本稿で行っている議論の方向性にも適したもののように思われる。しかしながら注意すべき点は、通話という方法が残る限り、聴覚障害者はそこに参加できないという状態も残り続けるということである。仮に、電話機等を用い、聴覚障害者と通信する際は、通話機能を用いずテキストメッセージでやりとりを行うことが聴者の義務でありなすべき配慮であると広く考えられるようになったとする。しかし聴覚障害者はいかにして電話の向こうの相手に自分が聞こえないことを知らせれば良いのだろうか。通常、見ず知らずの相手に、電話（通話）の前に何かしらの用件を知らせるということはあまりないだろう。重要な場面の通話等を除き、むしろ電話というのは事前準備を省いた即時のコミュニケーションと理解されているのではないだろうか。また、自分が聴覚障害者であることを相手に知らせ配慮を求めるという行為も、聴者のコミュニケーションへの参加障壁となるだろう。つまり、テキストメッセージ機能単体では、電話というコミュニケーションの代替とはなり得ず、加えて何かしらのアイデアが必要になってくると考え

られる。

翻って「みえる電話」について見ると、その新規性は、普段から電話を使う人（聴者）は、電話の向こうにいる相手が聴覚障害者であると電話に向かった時点では意識することなく、コミュニケーションが始まるという点にあるように思われる（通話開始後、「みえる電話」の機能について、アナウンスが流れる）。電話番号を入力する、あるいは電話着信があり、通話が始まったのちに、機械的に、そして機械を通じ、聴覚障害者—聴者コミュニケーションへの誘導が行われ、はじまる。いかに聴覚障害者であることを知らせるかという考察で見たような、別のコミュニケーションを開始するためのコミュニケーションという、聴者同士の通話場面では通常必要ない作業が、ここでは省かれる結果となっており、この点において、「みえる電話」サービスは、聴覚障害者に対する、通話機能の代替コミュニケーションとなり得る可能性を秘めているように思われる。一方で、音声からテキスト主体のやりとりに移行するというアイデアにおいても、その移行作業において、配慮を願い出なければならない場面が起こり得るのだが、聴覚障害者の側がテキスト入力で聴者側に知らせるのではなく、定型文や自動アナウンスが用意され、簡単な操作でそのことを伝えられるシステムとすれば、かなりの程度その作業に際しての抵抗感や心理的障壁のレベルを下げるができるだろう。機械的なシステムを用い人手が省かれることのメリットは、労力やコストが省かれることだけにあるのではなく、人や人の作業が介在してしまう際の煩わしさもまた、そこから無くなるということにもある。つまり、システム設計において作り出される、技術と人との境界面がどのようなものであるのか、という点が極めて重要になるということである。いうまでもなく、その設計作業には、実際に技術を利用する・必要としている人からの働きかけが欠かせないだろう。

以上、ここでは「みえる電話」サービスを見ながら、電話機等を用いた聴覚障害者—聴者コミュニケーションについての考察を行った。ここでも、障害者—非障害者コミュニケーションが、非障害者主体であり、かつ非障害者のコミュニケーション方法が据え置かれたものになっていることを確認することができた。そして前節のシンポジウムでの音声認識技術利用検討の場合よりも更に、問題は技術が存在しないことや未発達であるためでなく、目的を可能とする設計になっていないこと、つまり、技術を配置する人間の側の問題であることが確認できたと考える。

おわりに

以上のように、本稿では、障害者—非障害者コミュニケーションに着目し、とりわけ聴覚障害者—聴者コミュニケーションに範囲を絞った上で、音声認識技術の利用検討及び、実際に同技術が用いられているコミュニケーション場面からの考察を行った。その目的は、非障害者を多数派とする社会において人々が陥っている、既存のコミュニケーションが所与のものであり、負担を被って障害者へ情報を提供しているという考え方への再考を促すと同時に、実際にどのようなコミュニケーションの変容があり得るのかを例示することであった。十分に活用されていない技術とそれを用いたコミュニケーションという問題設定をした場合、まず個別の技術に着目することで実践に近い考察が行えると考え、ここでは音声認識技術を題材として選んだ。ここで用いた試案に登場する技術やその用い方は、ごく単純なものであり、専門家でなくとも考えつくことのできるものである。教育現場における音声認識技術利用の知が蓄積されはじめてるように、一つ一つは些細とも思われる知を提出しあい、集め、実践していくことが、より個別のコミュニケーション場面についても必要である。

注及び参考文献

- 1 河原 達也 2018「音声認識技術の変遷と最先端 ——深層学習による End-to-End モデル——」,『日本音響学会誌』74-7:381-386.
- 2 前掲 1:381.
- 3 二神 麗子ほか 2018「音声認識アプリを活用した ICT と人の協働による情報保障支援」,『群馬大学教育学部紀要』67:197-204.
- 4 前掲 3:197-198.
- 5 arsvi.com 2010「シンポジウム 「聴覚障害者の情報保障を考える」」, <http://www.arsvi.com/a/20100322b.htm>. (最終閲覧日 2020/12/13)
- 6 近藤 幸一ほか 2011「パネルディスカッション 障害者権利条約下におけるコミュニケーション支援の課題」,『生存学研究センター報告 :

聴覚障害者情報保障論——コミュニケーションを巡る技術・制度・思想の課題——』, 16: 223-249.

- 7 大野 ロベルト 2019「日本語のユニバーサルデザインをめぐって：視覚・聴覚に障害を持つ学生が共に学ぶために」,『日本社会事業大学研究紀要』65:101-114.
- 8 arsvi.com 2009「障害学会第6回大会」, <http://www.arsvi.com/ds/jsds2009.htm>. (最終閲覧日 2020/12/13)
- 9 アドバンスト・メディア社が開発・提供している技術で、UDトークなどに用いられている。
- 10 前掲6.
- 11 同前
- 12 坂本 徳仁ほか 2011「音声認識エンジンを用いた字幕化支援の現状と課題」,『生存学研究センター報告：聴覚障害者情報保障論——コミュニケーションを巡る技術・制度・思想の課題——』, 16: 144-156.
- 13 同前
- 14 河原 達也, 秋田 祐哉 2018「聴覚障害者のための講演・講義の音声認識による字幕付与」,『日本音響学会誌』74-3:156-162.
- 15 前掲14:157.
- 16 佐藤 庄衛 2018「聴覚障害者のための生放送番組への字幕付与システム」,『日本音響学会誌』74-3:163-170.
- 17 前掲16:165-166.
- 18 前掲16:165.
- 19 前掲16:163.
- 20 井部 千夫美 2019「米国における視聴覚障害者等向け放送の現状と課題」,『情報通信政策研究』3-1:91-106.
- 21 前掲20:94.
- 22 Pražák, Aleš et al. 2020 “Live TV subtitling through respeaking with remote cutting-edge technology” *Multimedia Tools and Applications*, 79:1203–1220.
- 23 前掲22:1206, 1212-1213.
- 24 アドバンスト・メディア社 2014-2020「よくあるご質問」, <https://www.advanced-media.co.jp/faq/proceeding>. ほか (最終閲覧日：2020/12/13)
- 25 アドバンスト・メディア社 2014-2020「モバイル・営業日報・業務報告についてのご質問」, <https://www.advanced-media.co.jp/faq/reporter>. (最終閲覧日：2020/12/13)
- 26 鳥羽 祐輔ほか 2017「情報保障システムにおける音声認識誤りの是正動機づけの検討」,『信学技報』116-453:19-24.
- 27 前掲26:22.
- 28 Arsvi.com 2020「障害学会第17回大会・2020」, <http://www.arsvi.com/ds/jsds2020.htm>. (最終閲覧日 2020/12/13)
- 29 前掲6.
- 30 前掲28.

Possibility of Transformation of Communication between People with Hearing Difficulties and Hearing Persons by Self-directed Action from Hearing Persons and Rearranging Technology: From Consideration of Use of Speech Recognition Technology

NAKAI Ryohei

Abstract:

Speech recognition technology has advanced technical innovation in recent years, in which one of the main purposes of using the technology is expected to be an effective tool to convert voice to text automatically for the people with hearing difficulties. However, most research on the technology use have sought how to apply it within a framework of communication between hearing persons, but not fully discussed the possible use beyond the framework and the possible communication between hearing persons and person with hearing disabilities. In such unequal situation of information access for people with disabilities, this paper focuses on how non-disabled people should work on problem-solving for the communication inequality. I examined how speech recognition technology services were used in different situations and how the technology can be applied to communication between people with hearing disabilities and the non-disabled. The result reveals that the lack of funds, technology or professionals is not a sole underlying problem with communication between them. Instead, the combination of different existing technologies is found to be a crucial factor to help people with hearing disabilities communicate with the non-disabled.

Keywords: communication for people with disabilities, Information Accessibility, social model of disability, people with hearing difficulties, speech recognition technology

聴者の主体的働きかけと技術の再配置による、 聴覚障害者—聴者コミュニケーション変容の可能性 ——音声認識技術の利用検討から——

中 井 良 平

要旨:

近年技術革新を経て認識制度を大幅に向上させた音声認識技術であるが、その主な利用目的の一つが、機械的な情報保障と目されている。同技術に関連する先行研究では、聴者のコミュニケーションという既存の枠組みに技術を適用させるという考え方が中心となっており、技術の活用可能性とコミュニケーションのあり方について十分に論じられているとは言い難い。本稿では、情報保障場面において、非障害者のコミュニケーションが据え置かれるという状態に着目し、音声認識技術の利用検討を通じ、聴覚障害者—聴者のコミュニケーションがどのように変容しうるかの考察を行った。その結果、技術革新や予算・人手の充実を待たずとも、コミュニケーション場面における既存の技術の配置を変え、その用い方に聴者が慣れることにより、聴者メインのコミュニケーションを聴覚障害者にとってより参加しやすいものとする事ができる可能性が多分にあることが明らかになった。