

第六章 日本で普及し始めた会議録作成分野の音声認識

一、地方議会を襲ったイノベーションの波

(一) 沼津市議会の衝撃

二〇〇四年六月定例会から静岡県沼津市議会が会議録の作成に音声認識を初めて導入したニューズは、速記界に強烈なインパクトを与えた。一九七七年には全国二百六十三市議会に五百六十三人の速記者がいたのが、二〇〇二年には四十六市議会に九十二人へと激減しており、速記録作成業務の外部化、速記離れが進んでいるからだった。とりわけ沼津市議会にはいわゆる速記士資格を持つ担当者がいなく、速記空白地帯で不特定話者の音声認識技術を利用して逐語発言記録を作成する点で、速記界の関心が高かった。

沼津市議会ではアドバンスト・メディアの議会議事録作成支援システム「アミボイス」を採用、一年間職員が復演方式で録音音声を復唱して会議録の作成に当たっていた。試行を終えた市議会では、本格運用に乗り出したものである。同市議会では二〇〇六年には会議録検索・閲覧システムを導入して市民がフリーワードで会議録を閲覧できるようにし、さらに二〇〇八年度からは会議の映像配信システムを導入して、我が国で初めて会議録と議場映像のリンクを自動化して市民へ情報開示している。社運をかけて取り組んだ同社では、速記録を作成する効率が二倍から四倍よくなり、外部委託費用の節減、職員の残業代を削減できる効果があると付言している。

(二) 北海道議会のチャレンジ

沼津市議会に続いて、都道府県レベルで初めて音声認識の採用に踏み切ったのは、ベテラン速記者を多数擁する北海道議会であった。技術革新の背景として、道議会速記室では厳しい財政状況から徹底したコスト削減と聖域なき施策の見直しが前提にあり、(一) 会議録作成分野においてもさらなる業務の効率化と経費削減が求められること、(二) 速記室は一九九六年以降八名体制であったが、二〇〇六年三月に一名退職し、二〇〇八年にはさらに三名が退職すること、(三) 速記職員の退職に伴い、補充採用を希望しても行財政改革の流れから認められるかは不確実なこと、(四) 衆参両速記者養成所が新規生徒募集を中止するとの情報を得ていたこと、及び民間速記学校などの養成状況からすると将来仮に速記者補充が認められても、その時点で新たな速記者確保が不確定であること、(五) 外注の活用による対応については北海道内において外注先の速記者が高齢化しており、今後民間速記者数の減少が想定されること、(六) 以上から、業務方法を改善しなければ早晚、担当する業務範囲に対応できなくなるのが確実なこと、という認識に達した。このため、二〇〇三年秋、五年後、十年後を見て、業務方法などの再検討を始めた。

沼津市議会が音声認識を本格運用開始したニューズを聞いた道議会速記室では運

用情報を収集、二〇〇四年六月にアドバンスト・メディアのデモンストレーションを見た。速記室主幹、山崎恵喜らは、沼津市議会を見学するとともに、同社の説明を受けて速記者が中心となって運用する見地から可能性を調査検討し、また道議会の本会議場、委員会室の環境調査などを行った。

その結果（一）速記者が管理するもとでシステムは十分実用に耐える、（二）音声認識導入により業務の効率化が図れ、減員になった状態でも対応が可能、（三）経費の削減にも結びつく、と判断し、二〇〇五年度当初予算に音声認識導入の経費を要求することに決めた。導入に当たっては沼津市議会のような試行期間をおかずにいきなり本格導入することとし、業務改善項目として（一）減員状態でも業務処理が可能とることから二〇〇六年退職者の補充を求めないこと、（二）記録作成の迅速化が図られ、一人当たり処理量が従来の二倍になり、外注を廃止しても内部処理で対応できるため外注を全廃すること、（三）原稿作成期間の短縮効果を利用して校正作業を行い、完全原稿の作成と、PDF印刷によって大幅な経費削減を図ること、を旨指した。

音声認識を導入する会議は本会議（年間百時間）、予算特別委員会（同百時間）、決算特別委員会（同四十時間）の規模を予定した。計画では、機種はアミボイスで、音声認識・文字化するレコーダー三台、音声と認識結果を蓄積するコントロール・サーバー三台（二〇一〇年から一台）、修正・編集作業をするライター七台の構成とした。本会議では発言は原則すべて登壇して発言し、予算・決算特別委員会では質問は専用の質問席、答弁は自席で行うことにした。所用経費は全体で約二千三百万円をかけ、五年リース契約で年間リース料は約四百六十万円（二〇一〇年から約三百三十万円）となっている。だれの声でも、自然な話し方で認識するのが特徴とされ、認識率は六〇〜九八％の範囲で、平均八〇〜八五％程度と見ているが、本会議と委員会、また発言者によってばらつきがあると想定した。二〇〇五年六月議会から運用が開始された。

音声認識エンジンでは、音響モデル、言語モデルが認識する際に重要な役割を果たす。認識率を高めるためには、まず音源の質を高くする必要があり道議会では音の大きさを一定に保つ音響コンプレッサーを導入した。最も重要なポイントは音響モデル、言語モデルの構築とメンテナンスであり、道議会では一九七〇年以降の会議録のテキストデータを同社へ貸与して言語モデルの辞書整備を周到に行わせた。それによって音声を文字化するときの認識精度を高めている。一方、可能な限りライターで行った修正データを辞書に反映することが大切なため年四回定期的に辞書更新を行っている。

二〇〇六年には「官公庁など公的機関、音声認識技術に注目。議事録作成効率化へ、来年度民間研究開発着手」（日本経済新聞）、「速記者は癖も読める、機械導入で激減だが」（朝日新聞）、「用なし、機械・・・方言に対応できず」（毎日新聞）、「法廷証言瞬時に文字化、最高裁開発へ。裁判員制度で活用」（毎日新聞）、「音声認識システム、

都議会も導入。議事録すんなり。ネックはヤジと話し言葉」（産経新聞）などと、マスコミを賑わせた。

北海道議会は、山崎ら速記者が構築した、都道府県レベルの議会における先進成功事例として評価されている。その理由としては、音声認識はあくまで記録作成の工程をサポートするものとしてとらえ、認識率を高めるために必要なメンテナンスをこまめに継続することを心がけ、一方自然な話を公文書記録として作成する上で必要な整文は人間が行うという原則に則り、機械を使いこなして高品質の公文書記録を作成するスタンスをとっているところにあると言える。

沼津、北海道の成功事例を追って、東京都福生市議会、福井県敦賀市議会、愛知県豊田市議会、大阪府東大阪市議会、岡山県真庭市議会、広島県庄原市議会、福岡県古賀市議会、長崎県南島原市議会、鹿児島県南さつま市議会など約四十議会がアミボイスを採用している。

（三）東京都議会の導入

アミボイスの三段跳びの最後は、東京都議会での採用であった。導入の経過をたどると、まず二〇〇一年、都議会は会議録作成システムの将来方向について検討を開始した。二〇〇六年には総務委員会の中で音声認識の試行システムを導入してテストに乗り出した。その間、沼津市議会等の導入効果、問題点を調査し、実施可能性を確認して、二〇〇七年から総務委員会において本格的な運用を始めた。段階的に適用範囲を拡大し、二〇〇九年には議院運営委員会、予算特別委員会を除いて全委員会へ拡大実施した。二〇一〇年五月には本会議へも導入を図っている。

その結果、長い間都議会から速記者へ特命で発注する形態から、都議会からアドバンスト・メディアへ特命契約で発注し、同社から速記会社へ発注するという階層構造に変更された。

都議会の場合も音源は従来からのアナログ音声回路を使用している。それをパソコンに取り込むことによってデジタル化し、音声認識サーバーと行き来する間に文字に変換され、その認識結果がネットワーク管理サーバーに保存され、速記会社の受信サーバーを介してエディターで会議録編集が行われる。エディターでは特殊なVテキスト形式で作業され、認識率が発言者ごとにばらつきがあるので、速記会社のサイドでジョブごとに音声認識結果を使用するか、音源からパソコンで書き起こすかを判断して記録原稿を作成する。

都議会の場合、認識率は試行時に七七%だったのが、本会議では九五%程度の高い水準にある一方、低い人では六〇%とばらつきが大きい。そこで、変換精度を高めるため、会期ごとの三か月に一度辞書の更新を行っている。文書の処理例を見ると、例えば認識結果が「都内、都外さんのは、都内産農産物」のように言い直して乱れた文

があったり、「都内産農産物の公社の対策に返す、冠水で調査研究について」のように意味不明の箇所があると、録音を確かめて正しい文章に訂正したり、整文の「言い直し」の処理基準を適用して直すといった作業を行っている。

本会議、委員会合わせて年間会議時間数はざっと本会議で四百四十時間ある。都議会は年間貸貸借料金としてアドバンスト・メディアへ約二千万円を支払い、速記会社へは記録業務（発注単価三千円）、修正業務（一万七千円）を実績に応じて支払うが、以前の契約に比べると六〇％程度に削減された。納期は会議が一時間以内に終了した場合は翌日夕方五時までに原稿を納め、都議会は速記録を議員へ一週間以内に提供するが、確定した内容は二十〜四十日後に印刷配布される。

導入の理由は（一）費用対効果を高めること、（二）今後の速記者確保の不安、にあった。導入条件は（一）会議録の質を絶対下げないこと、を大原則とし、その関係で導入当時からかわつてきた東京都議会局議事部議事課記録係主任、江村宇広は「速記者が持つ言葉を聞く能力は非常に高い。一般職の中には文字を好きな人、きらいな人がいて、もし、いやな人が配属されたら質の高い速記録は絶対にできない。だから、事務方が最終的な誤認識を修正する考えは基本的に持っていない」と導入決定時からの方え方を説明する。

（四）NECの活躍

NECは裁判市場だけでなく、二〇〇七年に議事録作成支援システム、二〇〇八年にはSaaS（必要に応じてインターネットから利用するソフトウェアサービス）型会議録作成支援システムなどを発売し、自動音声認識ニーズに込んでいる。音声認識分野のバイオニアとしてNECは多くの実用化技術を開発してきた。一九九三年にはパソコン音声認識ソフトを発売、翌年には大語彙音声認識装置「DS-1000」を出し、二〇〇七年十二月には議事録作成支援システム「ボイス・グラフィ」を発売した。同システムは、リアルタイムに音声認識するサーバーと複数管理するサーバー、さらに会議音声を再生し議事録を作成する編集ツールから構成され、声の特徴を自動学習する機能を国内で初めて採用したのが特徴であった。翌年春には愛知県議会、北海道美唄市議会が導入すると発表した。

ほかに導入事例として挙げられている地方議会は、北海道の北広島市議会、同新冠町議会、同浜中町議会、神奈川県大磯町議会、福井県小浜市議会、長野県箕輪町議会、島根県美郷町議会、広島県北広島町議会、福岡県飯塚市議会、同川崎町議会の十二を数える。

会議が終わったらすぐ議事録の草稿が完成するので、要約記録、逐語記録どちらへでも仕上げが行えるので、議会だけでなく、新聞社・テレビ局の記者会見のデータ処理、役員会、株主総会議事録、講演会記録、大学の講義録などに応用できるという。

(五) NTT東の議会用音声認識技術

NTT東は、自社開発した音声認識エンジンを用いた「会議録作成支援システムボイスエアー」を二〇一〇年十二月に発売した。これは、音声データから文書を自動的に作成できるシステムで、従来に比べ会議録作成時間が三六%も短縮できるという社内結果もある。音声・映像を再生しながら編集作業ができ、再生速度を自由に変えられるなど、操作性のよいのが特徴とされる。

システムの核となる音声認識は衆議院で採用された技術。同社では、新潟県佐渡市が音声認識システムの導入を検討しているとの情報を得て、二〇一〇年十一月に本システムを提案した。翌十二月に開いた模擬実験の臨時市議会を録音して直ちに認識処理を実施した。その日のうちに仕上げた会議録を市幹部や市議会議員へ配布したところ、速報性、認識精度について高い評価を得た。その後、二〇一一年三月に実証実験を行い、また他システムとの比較検討を経て採用が決まった。

二、衆議院の音声認識開発物語

(一) 国会の威信をかけてイノベーションに挑戦

一九一八年、国会は第一次世界大戦による好景気に伴って早晚優秀な速記者を充足できなくなるのではないかという事情から、貴族院に速記者練習所、衆議院に速記者養成所を新設し、五月二十日から衆議院速記者養成所では新井田次郎が初代教官となって授業を始めた。太平洋戦争、高度成長、オイルショックを経験して時代は大きく変わり、日本経済はバブル崩壊後衰退を続け、行政改革の波が国、地方に及んでいた。そのような時代の流れの末、衆議院は二〇〇六年十二月、速記者養成所を廃止し、将来に向けて新会議録作成システムを開発、導入することになった。

さかのぼって二〇〇四年の議院運営委員会で速記者養成所の新規生徒募集中止が決定されていたので、事務方で同年十月に検討プロジェクトチーム（座長・事務次長鬼塚誠）を編成して、今後の記録方法についての調査、検討を始めていた。

今後の方法としては、現在参議院が採用しているパソコンによる聞き打ち方法をとるか、新しい技術を導入するかを検討した。新技術では、日本語機械速記の電子化を図るはやとくん（ソクタイプ）が本命候補とされたが、最終的にはまだ実用化されていない不特定話者による自然言語の音声認識技術を開発する方針に傾いた。二〇〇五年、自動音声認識技術研究で先導的な立場にある京都大学教授、河原達也へ技術顧問を委嘱し、衆議院における会議録作成システムの構想立案に乗り出した。

衆議院では、音声タイプライターの一号機開発からこの分野の技術ノウハウを蓄積している日本電気に新会議録作成シス



テムのプロトタイプ開発を発注した。でき上がったプロトタイプを一年間試行した。速記者は、議場に赴かないで、自室で机上のパソコンを用いて自動音声認識技術で逐語速記録の原稿を作成するというものであった。衆議院速記者八名によるプロトタイプの評価では、現行速記書き起こし方法に比べて音声認識のほうが編集時間を短縮できることがわかった。その場合の認識率は七五%で、約半数が音声認識に否定的な意見を述べたという。しかし、認識率が八〇%以上になると、否定的な意見は皆無であった。大学生を対象に行った実験では認識率七五%を境に音声認識が好まれる閾値であると見ていたことが一応裏づけられた。

その検証を三菱総合研究所へ発注し、評価を行った。その結果、日本語の音声認識システムの開発は可能と見て、衆議院では二〇〇八年に入札を実施したところ、日本電気に競り勝って、東日本電信電話（NTT東）が落札することに成功した。衆議院は新会議録作成システム開発に際して辞書が最も重要な構成要素であると位置づけていたが、その意向を汲んで辞書分離を提示したNTT東が受注したものである。契約では、京大が開発した音響モデル、単語辞書、言語モデルを衆議院にライセンス供与した上で、ベンダーが認識エンジンと統合する形であった。

中核となる辞書の作成に関して、河原に、十年以上分の衆議院の会議録データと国会テレビ（一九九二年から開始）などからの音声データを提供して辞書構築を依頼した。河原は「私もこれまで国会審議を忠実に書き起こし、音声と対応づけたコーパス『衆議院審議音声コーパス』を構築してきた。これは衆議院の主要な委員会と本会議をカバーしており、合計二百時間、二百四十万単語の規模となっている。このコーパスをベースとして音響モデル、単語辞書、言語モデルを国会審議で用いられる話し言葉音声に対応できるように改善した」と説明する。

加えて、河原は会議録データと実際に話されている内容の違いとを分析して、速記者が行っている言語処理を継承した変換モデルを作成した。例えば意味を持たなくなった無機能語を取り除くフィルターの処理や、「い抜き言葉」と呼ばれる「考えてる」を「考えている」と「い」の文字を加える処理など、公式の記録文書を作成するための整文基準が織り込まれた。これらの配慮が総合的に変換効率の改善に貢献、当初八〇%程度と見られていた効率が運用後には平均八五%、本会議で九〇%を上回り、委員会でも八〇〜八七%の実績を挙げており、統計的にデータを集めることによってかなりの予測が実現できるようになったという。河原は「この性能は商用ソフトウェアでは実現できない」としており、世界最先端のプロジェクトが学官連携で成功した。

開発費用二億円を投じた新会議録作成システムがNTT東の技術陣によって完成され、二〇一〇年から試行運用された。最初に試用されたのは、五月十九日の外務委員会だとされるが、対象会議が次第にふやされていった。ネット報道によれば一年間の試行を経て二〇一一年四月十二日の予算委員会から本格運用が始まった。二〇一一年五月十三日付の朝日新聞、日本経済新聞、京都新聞などは京大の音声認識技術が衆

議院で導入されたと一斉に報道した。

(二) 会議録作成システムの運用

日本の立法府は二院制をとり、下院に当たる衆議院では本会議のほか、十五の常任委員会と六〇八の特別委員会を持っている。衆議院の職員は千七百名で、記録部には百六十名が在籍しており、その多くは速記者である。

新しい作業方法によれば、前日午後遅く翌日の会議予定が決まって、速記者に原稿作成の担当任務が与えられる。担当速記者は自分のパソコンに予定を入れておくと、コンピュータシステムは会議を自動的に五分ごとのセグメントに区切る。会議が始まると、自動的に五分刻みにされた音声データが保存されるとともに、音声認識エンジンへ送られ、認識を終えたデータは時間順に担当者へ配られる。担当速記者は音声データからテキストデータを校正し、二回校正して速記原稿を完成する。

ただ、議事録を作成する際、議場にいなければ得られない情報が幾つかある。例えば委員長の指示、自席発言、拍手、うなづきなど非言語コミュニケーションがあり、議場のそのような情報を担当速記者がわからなければ正確な会議録を作成する上で支障が生じる恐れがある。この種の課題を解決するために、記録部では新しく「臨場者」(ロガー) という職務を設けた。主に経験のある速記者を充てるロガーは議場の中心で会議の全体が見え、よく聞こえる場所に席を取り、発言順に発言者名、答弁要旨などを、五分間分ずつA4判用紙に記入して、担当速記者へ渡すことが役割である。二〇一一年三月には試行の最終段階にあった。

ここで重要なことは、議場での速記者は自動音声認識システムに置き換えられたが、会議録全体の作成業務の流れは変更されたところはない。担当速記者が完成した原稿は二十年以上の経験があるベテラン校閲者へ送られる。会議が深夜になったらすると、原稿は翌日提出されることもあるが、基本的には当日中に仕上げられ、校閲者もその日のうちに会議録を完成する。ここでいう原稿は公式のものではなく、プリントかPDFで議員等へ配布される。最終的に会議録原稿が完成されると、印刷局へ送られ、電子データがインターネットで公開されている。

委員会の審議は衆議院のホームページからリアルタイム映像配信されている。審議が終わると、その日のうちにビデオライブラリに保存されて、閲覧することができる。ただ、衆議院ですべての会議が自動音声認識にかわったわけではない。本会議、予算委員会、議院運営委員会の三会議は、依然として速記によって逐語記録が作成されている。これらの会議は最も重要な会議であり、会議録の配布を迅速に行う必要があるからである。室長の猿谷豊は、その理由を「議場に出て速記したほうが反訳原稿の仕上がり非常に早い」とする。議院運営委員会は応接サロンの部屋を使って会議するので、規格の録音設備がないため手書き速記で継続している。それから、ビデオ記録

が禁止されている秘密会では速記で記録をとるようになっていた。

衆議院において自動音声認識技術が順調に進んでいるのには、幾つかの理由がある。まずその可能性の限界を直視することであった。技術的に進んでいても、音声認識は一〇〇%正しい変換をすることはできない。文書として必要な句読点、改行ができない、様式を整えることができない。発言者名も出ない。このような課題を解決するため、変換では辞書に手をかけてよくしていくこと、後者の課題は速記者が新しい手法を利用して経験ノウハウを発揮する点にあると見た。

変換精度を上げるために、記録部専用回線による非圧縮のデジタル音源とした。音声はステレオ2チャンネルで、重複発言に耐えるようにマイクは発言席と答弁席（委員長席）に分けた。また国会CATV中継のテレビ映像を記録部のサーバーで録画しており、音声はモノラルながら、速記者が自席で映像と音声を聞くことができる環境とした。さらに、デジタル音源設備がない委員会等については、非圧縮のリニアPCMで録音している。

音声認識エンジンは、日本電気が自社開発技術による限りなくリアルタイムに表示できるバッチ処理で、自社が構築する辞書を使用することが条件であった。それに対して、NTT東はバッチ処理であるが、衆議院の希望にかなって、京大河原が構築する辞書を組み込む点だけで差異があった。技術的には両社とも合格であったが、河原が長年データを蓄積して認識率の向上を研究してきた先行研究が衆議院の求めるところと合致したと言える。音声認識の単位は、音声タイプライターの時代は一音ずつに認識していたが、次に単語単位に進化した。現在は、前後の単語とかかわりを見極める三単語で処理するようになり、速度、精度両面で十分な実用性を見極めていく。

辞書充実には、担当速記者が原稿を作成する作業の中で、音声データと文字表記のデータを固有名詞や初出語、未知語登録しておく方途もとられた。また、以前は発言に超忠実なテキストがないとできなかったが、近年は会議録テキストと音声データがあれば会議録の整形もできるように配慮された。

そのツールとして衆議院が誇るのには、独自設計のエディター「速手」（はやて）である。エディターは、一般に文字情報の入力、削除、コピー、貼りつけ、検索、置換、整形などの機能を備えている。河原は「音声認識とともに重要な検討事項は、エディターだった。認識誤りを訂正し、会議録としてふさわしい文書体に整形する作業を効率よく行う上で、極めて重要な要素だ。エディターは、元の音声を時刻指定や文字の上のカーソル指定や発話単位で参照できる機能を提供するとともに、音声再生を速くしたり遅くしたりすることも必要だ。衆議院の速記者が中心に検討を行った結果、一発話ごとに修正する方式のラインエディターではなく、ワープロ型のインターフェースのスクリーンエディターが、編集作業に集中できる点で優れているとの結論になった」と述懐しており、ラインエディターしか用意していなかったベンダーと最も見解の違いがあった点である。

作業者の視点から操作性を検討して、校正画面上のカーソル位置は最下行から上へ四行目に来るように設計されている。担当速記者は認識結果を読みながら再生音声と校正し、間違いや整文処理の対象を発見すると適切な処理を施す。文書として形を整える作業では、臨場者が作成したメモを参照しながら、発言者名を加え、不要な部分を削除、段落を変えるなどの作業がある。例えば、「おはようございます」というようなあいさつは会議録に記載する必要はないので、省く。それから、「えー」「あー」や、本来意味を持っていた言葉が発言者の話癖になっていて、本来の意味を示さなくなり、余り多くて文書の意味を汲みにくいような場合には取り去る。新聞社等で記者ワープロに組み込まれている校正ソフトのように、速手のエディターから入力してもし間違いがあると、赤色文字に色を変えて注意を喚起するほか、外字を用いると青字に表示される仕組みが組み込まれた。

二〇一一年の東北大震災のあとの国会では地震名、専門語が使用されたとき、音声認識ミスが出たが、それらの言葉は速やかに辞書登録してどんどん知恵を与えていく。「原子力発電所を遅い」という認識変換が発見されると、すぐさま「原子力発電所を襲い」を優先して選択するように辞書を鍛える作業が速記者全員の共同作業で行われるため、辞書に未登録な単語、コンピュータが初めて受け取った音、文字列、文脈から間違った文字表記などは人海戦術で教え込む仕組みとされた。ちなみに、辞書ビルディングは、(一)新聞情報、(二)用語集、(三)活字資料、(四)配付資料、(五)法令略語、のカテゴリから作業する速記者自身が補強していく。要するに、現場の努力で自動音声認識システムを学習させながら進化させていく決意が秘められている。

開発費用二億円を投じた新会議録作成システムがNTT東の技術陣によって完成され、二〇一〇年から試行運用された。最初に試用されたのは、五月十九日の外務委員会だとされるが、対象会議が次第にふやされていった。ネット報道によれば一年間の試行を経て二〇一一年四月十二日の予算委員会から本格運用が始まった。

三、参議院の現実的な選択

参議院は、二〇〇五年十月二十二日の議院運営委員会で、速記者養成所の生徒の新規募集停止を決定した。すでに募集活動を始めていた衆議院は急遽募集中止を決定して参議院に追従した。これによって衆議院から検討を要請されていた速記者養成所の統合問題は消滅することになった。貴族院から参議院に変わった一九四七年、参議院は四十六名の速記者で発足し、今日九十名弱の速記者と三十名強の校閲者、会議録発行のために後方支援者が四十名弱の陣容となっている。情報化社会の進展に向けて、国家公務員定数の削減要請のもとで、会議録の早期発行、データベース公開など、新規の情報開示サービスの要請があり、募集停止後、新速記システムの検討を進

めることになった。

検討で取り上げられたのは機械速記と自動音声認識システムであった。機械速記ではソクタイプを取り上げたが、検証には遠藤が開発中のCAT（電算支援反訳）を対象にした。最高裁はCAT開発では可能性を探究したレベルで終わったため、はやとくんの会メンバーが私的に使用しているシステムに着目したものである。二〇〇二年には実証試験を実施し、議会発言への対応可能性、反訳精度、反訳能率、習得の難易度、教育機関の有無、速記機械と速記ソフトの確保、ソフトの機能改善の難易度、外国製速記機械のメンテナンス上の課題などを検討し、「以上のことから、速記官仕様の機械速記を記録部に導入することは、解決すべき課題や解決困難な問題も多く、総合的見地から消極に解さざるを得ないとの結論に達した」と結んだ。

自動音声認識技術に関しては、参議院は国会の会議録は憲法あるいは議院規則で成が義務づけられており、「速記法によって、すべての議事を記載しなければならぬ」とされている点を挙げて、慎重な立場をとった。特に認識率については「九〇、九五％が限界ではないかと考えられる」として、記録部が求める精度一〇〇％に対して現行の作成方法が平均九九・五％である状況からみて、機械速記、自動音声認識はいずれも要求を満たせない上、解決すべき課題や問題点が多々あるとして二〇〇三年から新たな作成方法をさらに調査すると結論づけた。

新速記システムの検討では、二段階に分けて予備調査を実施した。そのシステムの前提は、日本放送協会が開発した「映像・音声変速再生技術」であった。第一段階では速記者がシステムに慣れることに主眼が置かれた。第二段階では模擬的に実践形式で実施、作成速度、精度のデータを詳細にとった。

参議院は機械速記、自動音声認識、話速変換システムとを比較した。リアルタイムに配信される映像、音声を視聴しながらパソコンのキーボードから即時入力する新手法が機械速記、自動音声認識と比べ、(一)これまで検討したシステムに比べ、ハード、ソフトとも不十分である割には、原稿の完成度が高く、リスクも極めて少なかったこと、(二)現在の速記者の業務形態を大きく変更することなく、また二人一組から一人一組体制に円滑に移行できる等のメリットがあると考えられること、(三)速記者の負担を一定程度軽減できるとともに、余剰時間が生じる場合には審議内容の精査等に有効活用できることから、一層の速報化を実現しつつ、従来の精度も維持できる等の効果が期待されることと、現状を踏まえて移行できる妥当な選択と判断した。また議事は速記法で記録するという規則については、音声を即時入力できることや第一次原稿を機械速記と同等の時間で作成可能だから、「これも速記と考える」と速記の解釈を示して、今後は「新速記システムとしてさらに調査、研究を進める」とした。

記録部ほぼ全員が参加して全般的な検討を行い、二〇〇四年には概算要求に向けて新速記システムのプロトタイプ構築を目指した。議場に速記者が出ないで、再生速度を遅くしたビデオ、音声を聞きながら、パソコンのキーボードから入力する仕事は参

議院に定着した。ビデオ録画は速記者全員が共有し、当速記者へは五分刻みで配信して直打ちしていく。ただし、主任記録者が議場に出て発言の出だしをデジタルペンで書き、入力者へ送る。

一般職を対象に整文等の技術を継承する人材育成も始められている。

四、最高裁のとった選択

一九七二年、最高裁はCAT実験を開始するとともに、録音筆稿方式の実験を東京地裁へ依頼した。それを受けて、書記官協議会が中心になって東京録音反訳委員会設置のための推進準備委員会が発足し、会則案、反訳従事者名簿作成など録音反訳者の組織指導の準備を始めた。大蔵省はその実験費一千万円を認め、原本分の予算を措置した。ソクタイプ符号自動反訳実験と並行して進んでいく。録音筆稿方式は、もともと一九七〇年、東京地裁で考え出され、民事部の裁判官主任会議の議決を経て実施に移っていた。民事三十五部で実験を開始したので、俗に「三十五部方式」と呼ばれてきた。当初、最高裁は筆稿業者が作成する裁判所の調書の原本分まで訴訟当事者に負担させるという条件であった。ところが、この方式を続けていると、この方式が裁判手続きの合理化政策に合致すると見た最高裁は一九七一年に至って設備費及び原本分の作成費用等について予算化し、三十五部と十二部を実験部に指定して、実験の名目で普及していくことになる。長い問受け皿を司法協会が担ってきたが、二〇一〇年には約二億七千万円の市場へ成長した。二〇〇九年に競争入札方式に変更されているが、二〇一一年の入札の結果、料金の低下が大きく、司法協会は消え、売り上げの多くを依存してきた民間業者一社も落札できなかった。年間三万時間の大市场だが、安定受注の保証がないので、経営リスクが非常に高いことが明らかになった。

近年、録音テープを支給して裁判速記調書原稿を作成する形態から、裁判員裁判を反映してNECの自動音声認識ソフトを使ってテキストファイル化したデータを基に録音反訳業務を委託するという形態へ変化した。民間業者へ送られてくるのは、音声データとNECフォーマットのテキストデータ、それに関連資料と立ち会いメモ形式である。認識結果と音声を照合して校正し、裁判調書の体裁に整える作業を行う。しかし、発注の仕様書には自動音声認識の最低認識率の保証がなくて、人によって全く認識されていない場合でも成果品を納品しなければならぬこともある。納期は二十四時間と四十八時間のいずれかとされ、審理のスピーディな進行を支えている。

このシステムは、二〇〇九年裁判員裁判向け音声認識システムとしてNECが受注していたもので、二〇〇六年から三年間にわたって最高裁の協力を得て研究開発していた成果をベースに開発された。認識率は二〇〇六年度の開始時には六〇%を切っていたが、新規に開発した技術、モデルを投入した。大体八〇%あれば、裁判の音声、映像の検索、頭出し再生機能として活用する際に実用に耐えらるると考えられ、二〇〇八

年後半には一定程度の認識率を実現できており、実用性を徐々に高めていると言われている。最高裁が音声認識技術の利用に投じた費用は衆議院を大きく上回ると伝えられている。

N E C技術陣が新たに開発した技術の一つに、重なり発言などに対応するマルチチャンネル音声検出技術がある。法廷環境には発言者の主な四種類の役割ごとに一つのチャンネルを割り当て、四チャンネル同時に並行して認識する方式を採用した。同時発言に対しては二つの音声と同時に並行処理させ、一人の発声が二つ以上のマイクにも回り込んでいる場合には一つの処理だけを動作させるようにした点が特徴である。また、高速処理を実現するための新技術としてリアルタイム教師なし話者適応技術や役割別音響／言語モデル、さらには関西語への言語モデル変換技術、句読点自動挿入、フィラー除去などの技術を投入している。

こうした成果を見て、検察庁も取り調べの可視化のために自動音声認識の採用を検討していると言われる。

五、放送字幕への音声認識の応用

自動音声認識技術はリアルタイム性に優れている。NHKの場合、地上波デジタルテレビ放送につけている字幕の比率は二〇一〇年実績で五六・二%であり、字幕をつけられない番組を除くと六二・二%となっている。それを二〇一七年には字幕対象の番組すべてに字幕をつけようという目標が与えられていた。

字幕には番組を編集してから音声を書き起こして字幕を作成し、画面に入れていく方法と、ニュースや野球中継に生で入れていく二つの種類がある。NHKの字幕放送は、一九八三年十月、朝の連続テレビ小説「おしん」に実験的につけたのが始まりである。そして一九八五年十一月、朝の連続テレビ小説「いちばん太鼓」から本放送を開始した。

N H Kが音声認識の研究に手を染めたのは一九六九年だった。本格的に生字幕制作システムの研究に乗り出したのは、一九九六年であった。ニュースはアナウンサーが朗読する部分が多いので、ニュース用の音声データベースを構築するとともに、音声認識手法の改良に努めた。一九九九年には認識率は九七%に向上し、二〇〇〇年三月にはスタジオアナウンサーの原稿読みに限定したダイレクト方式の音声認識技術で、「ニュース7」で日本初の生字幕放送を行った。ただ、認識エラーを訂正するため、タッチパネルで誤りを発見するオペレーターとキーボードから訂正するオペレーターの二人一組の二組を要した。また、アナウンサーの読み上げ部分に限定していたため二〇〇六年にその運用を中断し、スピードワーププロに移行している。

ほかの番組では、スポーツ中継、歌番組などについても生字幕の拡充を検討、欧米で確立されているリスピーク（復演口述）方式を適用して二〇〇一年末の「紅白歌合

戦」を手始めに「オリンピック」、「大相撲」などの実況中継へ生字幕を拡大してきた。こうしてリスピーク方式の経験を高めていくと、音声認識装置に登録できる語彙が四万語から十万語に拡張したことによって、音声認識が使える番組が年々増加している。二〇一〇年四月から「スタジオパークでこんにちは」での運用を始め、情報・バラエティ番組へ拡大しつつある。

さらにさきに運用を中断したダイレクト方式に、リスピーク方式を切りかえ式で組んだハイブリッド方式を試作し、街頭で一般人のインタビュー区間をリスピークに切りかえて音声認識する方式を実験している。

ここで、現在NHKが使い分けているリアルタイム字幕入力法を見てみよう。

【歌謡番組・情報番組で利用】通常のキーボードを利用する連係入力方式で、複数の入力者が短い時間を担当してリレー式に文字を入力する方式

【ニュースなど報道番組で利用】電子速記法のステノワードを用いて入力するやり方で、入力者とスコピスト（校正者）のペアが複数組みで短い時間を分担して高速発話に遅れないように入力していく方式

【大リーグと二〇〇〇年～二〇〇六年のニュースで利用】番組音声を直接認識して文字に変換するダイレクト方式

【相撲や野球などのスポーツと情報番組で利用】字幕専用のアナウンサーが復唱した音声を文字に変換するリスピーク方式

生字幕の課題としては（一）正確さ、（二）リアルタイム性、（三）効率的な確認、訂正、（四）事前準備の容易さをNHK放送技術研究所の今井亨は挙げている。

どのような方式でテレビに生字幕を付与するかは別として、つきまといて離れない課題に表示の遅延がある。テレビ朝日データビジョンによると、NHKニュース（二〇〇六年十一月八日放送のニュース7からサンプリング）における字幕提示の遅れは平均八・三秒、最小遅延時間五秒、最大遅延時間十六秒であったという。

もし、発話速度が平均的な三百字／一分間と仮定すると、八秒の遅れは約四十文字となり、音声にかなり遅れてしまっている。十六秒遅れだと八十文字になり、話の流れについていけない。その原因の多くは、エラーの発見、訂正にあり、このスコピング作業の革新がない限り生字幕の視聴者満足度を改善する方法はないであろう。