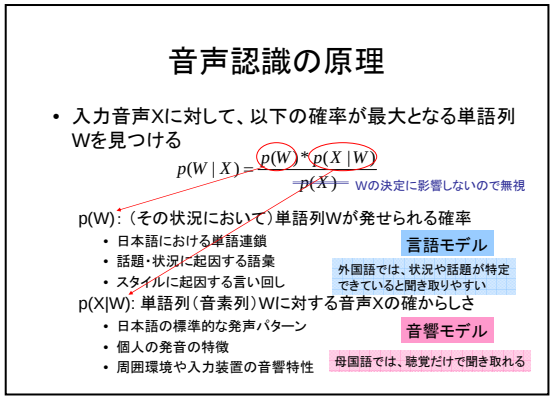
第58回全国議事記録議事運営事務研修会

音声認識技術の現状と会議録作成への適用可能性(2)

京都大学教授　学術情報メディアセンター　河原　達也氏

・音声認識の原理



いきなり式がついて恐縮なんですが、音声認識の原理を一言で言うとこの式になってきます。数学的にというか、記号を使っていいますと、入力音声（X）に対して、この確率が最大となる単語（W）を見つける。音声が入ってきたときに、これは「ハロー」と言っているのか、「はい」と言っているのか、いろいろ考えながら聞いていると思うのですが、その確率が一番高いと思うもの、こう言っているのだろうという確信の高いものを最終的に認識結果とします。

言語モデル、音響モデルという二つの確率の項があります。音響モデルは単語列、あるいは「あ」「い」という音素に対して、その音声とはこういうものであるということ。子供のときから言葉を聞いていると、自然に耳ができてくるのと同じです。個人の発音の特徴も、ある程度入ってきますし、周囲の環境を多少考慮してつくります。

それに対して言語モデル。ネイティブの母国語話者は、音響モデルだけでほとんど聞き取れるようになるのですけれども、特定の状況において、この単語列が発声される確率、日本語でしたら例えば「私」の次には「は」が来るとか「の」が来るという文法的な制約もございますし、もうちょっと言いますと、例えば政治の話をしていたら政治の単語が出てくる、経済の話をしていたら「株」「為替」という単語が出てきやすいといった、話題とか状況に応じて出てくる語彙とか表現がある程度決まってくることを考慮してつくります。

それから、スタイル。議会とか講演でしたら、「ございます」といった丁寧な言い回しがされるであろうといったことです。このような状況を想定して、そこでどういう表現が使われているかを予測すると、聞き取りやすいということです。英語を聞く場合は非常にわかりやすいと思うのですけれども、北朝鮮のニュースをしていたら「アトミック」「キ

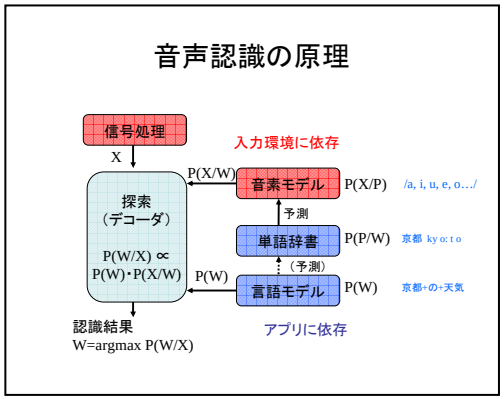
ム・ジョンイル」といった単語が出てくるということが何となく予測できますから、聞き取りやすくなってくるわけです。外国語を聞く場合には、非常に重要なかぎになっているのです。外国語では、こちらの言語モデルがないと聞きにくいし、あると聞きやすくなります。

・音声認識の基本要素

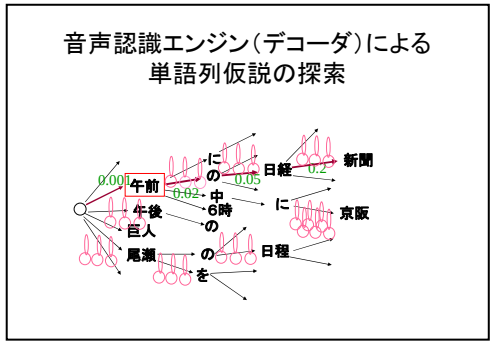
このシステム、機械をつくるときは、さっき申しましたように、日本語の言語学的な知見、「私」の次には「は」「の」が来やすいとか、話題とか状況をある程度モデル化するとか、会議でしたら、「何々に関しては何々でございます」といった一定の表現、言い回しを機械に学習させます。こういう表現がよく来るんだというスタイルを機械に覚えさせるわけです。

会議と電話の会話では、話題や言い回しが全然違いますから、それぞれタスクドメインごとにつくらないといけないということになります。

一方、音響モデルの方は、純粋に「あ」とか「い」という声をたくさん集めてきて、性別、環境ごとにパターンをつくります。カーナビと電話では当然環境が違いますから、それぞれ音のパターンを覚えさせてつくります。



これがブロック図にしたものです。音声に周波数分析みたいなことをします。それに対して、先ほど申しました音響・音素モデルと言語モデルは、全部確率的な枠組みでやっていますので、全部尤度に落ちます。天気の案内をするアプリケーションを考えるなら、言語モデルでは「京都の天気」「あしたの空模様」といった表現が来やすいというのを覚えさせて、「京都」は「KYOTO」というスペルも覚えさせ、「K」「T」はこういう音のパターンだということを覚えさせて、それぞれを総合して認識します。最終的に「京都の天気」と言った確率が一番高ければ、それを認識結果にするのです。



これは、その過程をイラストにしたものです。外国語や英語を聞いているときに、ああ言っているのか、こう言っているのか考えながら聞くことがありますが、それを機械でもいろいろやっているのです。

これは我々の宣伝になるのですが、私どもの京都大学が中心になって、フリーの音声認識ソフト「Ｊｕｌｉｕｓ」を十年ぐらい前からつくっております。このウエブサイトは、「音声認識」で検索すると五番目以内には出てくると思うのですけれども、「Ｊｕｌｉｕｓ」と入れると、多分一番に出てくると思います。

この特徴は、商用のソフトと違って、フリーソフトということ。タダということもあるのですが、ソースコードも開示しております。パソコンで言うとウインドウズとリナックスというのがありますが、リナックスの方は中身が全部見えるようになっている。それに近いものでありまして、一般の方にとって必ずしも使いやすいわけではありませんけれども、システムを開発する方には非常によく使っていただいています。

それから、市販のものと違って、リナックスでもウインドウズでも動くようになっていますし、最近は某メーカーの方のマイコンにも移植してもらっています。また、日本語だけではなくて、英語やいろんな言語に持っていけます。

ただ、公開しておりますバージョンは、基本的にはディクテーションのモデルでありまして、かなりゆっくり丁寧に発声しないと認識できません。

では、なぜそれしか公開しないのかと言いますと、先ほども言っていますように、それ以外のものは全部オーダーメードになってきますので、議会なら議会用にコツコツデータを集めてつくらないといけないという過程がどうしても要ります。万能で何でも認識できるというレベルでは、丁寧に発声することを前提にしなければいけません。

オーダーメードで開発するときにかぎとなりますのがコーパスデータです。さっき音響モデルと言語モデルという二つのモデルがあると言いましたけれども、それぞれ統計量を学習させます。人間の場合でも、たくさん子供のころから言葉を聞いていると、だんだんわかるようになってきたり、英語でも、何十時間、何百時間とヒアリングの経験を積むと、英語のパターンが身についてきますけれども、それに相当することをやっております。

具体的には、音ですと少なくとも数十時間、できれば数百時間聞く。当然、一人の人の声だけ聞いていてもいけません。いろいろな話し方の人のを聞かないと、聞き取りはできませんので、目安としては数百人が必要と言われています。

言語モデルの方は、数千万語ぐらい。数千万と言われてもピンと来ないと思いますけれども、新聞記事で言うと十年分。あるいは今、日本じゅうに数億あると言われていますウエブページを全部集めてきてつくるということをやっています。

ただ、それらは基本的には書き言葉です。どうしても議会のような話し言葉にはそのままでは適用できません。ですから、議会の会議録といったものをベースにつくっています。

・話し言葉音声認識の問題

では、具体的にどうつくっているかというお話です。

これは、話し言葉と書き言葉、読み上げ音声と実際の自然な音声でどういう違いか、これも体感していただくために、音を出してみます。今度は日本語です。

サンプル一　また、政治に金がかかるということ自体がわからない。

サンプル二　このお話、幹事の方からいただいたときも、そのやりとりの中でも実はあったのですが……

こんな感じで、最初のように丁寧に発声すると九五％行くのですが、後のように、本当にそのまましゃべっているのを生で入れると六五％になります。

二つを聞き比べていただきますと、どういう違いがあるかといいますと、話し方の、特にスピードが後の方が速いということもありますし、話し言葉の特徴として、速くなったりゆっくりになったりします。考えているとゆっくりになりますし、スラスラ出てくれば速くなるといった大きな変動があります。音声認識システムにとっては結構クリティカルで、一定速度で発声しないと認識できないという特徴があります。

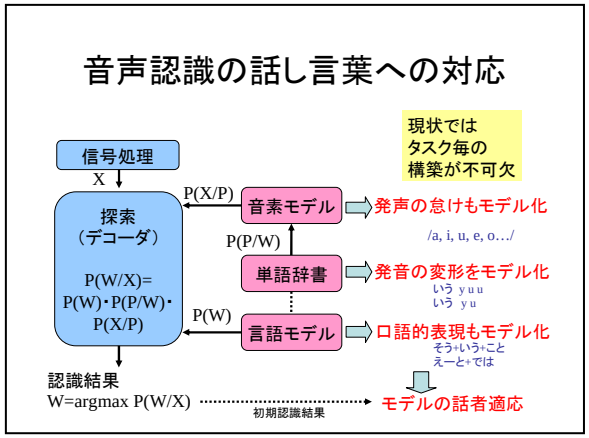
あとは、発声の怠けとか口語表現であるとか、言いよどみといった現象があります。

また、話し言葉の特徴として、「何とかですが、何とかでして……」と、一分でも二分でも話す方がおられますけれども、発話の区切りがよくわからないといった現象もあります。

もう一つ、一番の特徴として、話し言葉の場合、個人差が大きいということがあります。つまり、表現にしても、発音にしても、クセというか方言も含めまして、標準語を意識してしゃべればそろうのですけれども、地の言葉になると差が大きくなりますので、できることならば、人ごとに適応していくことが望ましいと言われています。基本的には、音響モデル、言語モデルと全部話し言葉用にカスタマイズしてつくります。

この図は、さっきと同じブロックダイアグラムです。発声の怠けや発音の変形、話し言葉の口語的な表現をモデル化します。それから、話者ごとに適応します。

現状では、これらをタスクごとにつくる必要があります。



・議会向け音声認識モジュールの構成

これは、今、京都大学で研究開発しています衆議院向けの音声認識モジュールの構成です。現状はＣＳＪと書いておりますが、学会講演を二百時間ぐらい集めたものと、衆議院の会議録をベースにつくっています。現在、衆議院さんの方から百五十時間ぐらいの実際の委員会等のデータを提供していただいておりまして、それをベースにこれらをつくっています。現状では、予算委員会で八〇％ぐらいの認識率ですが、目標としては、一年以内にこれを八五～九〇にしたい。これは確たる成算があるわけではありませんけれども、現状よりは幾らかよくなるだろうという見通しです。

衆議院向け
音声認識モジュールの構成（京大）

	現状	将来（研究開発中）
音素モデル	CSJ	CSJ+審議コーパス
単語辞書	会議録+CSJ	会議録+審議コーパス
言語モデル	会議録+CSJ	会議録+審議コーパス
認識率	80%（予算委）	85-90%??

CSJ：日本語話し言葉コーパス（学会講演；228時間）
審議コーパス：　衆議院審議の忠実な書き起こし；150時間

・衆議院審議コーパスの例

これは実際の衆議院のコーパスの例です。黒い字で書いてありますのが会議録でありまして、これは皆さんの方が身近でよくおわかりかと思います。

実際の発話を忠実に書き起こしますと、「えー」「そのー」が随分入っております。赤い部分がフィラーです。「ですね」も会議録では落とされています。

それから、「だけど」が「だけれども」になったり、「なるんで」が「なるので」という文書体に整形することもなされています。

「いただいてる」を「いただいている」のように、「い」を補完するなど、実際の忠実な話し言葉と会議録でも随分ギャップがあります。衆議院の予算委員会の例で数えてみますと、単語のうち一二％ぐらいが違っていると観測されています。

ですから、会議録は実際たくさんあるのですけれども、本当に話している言葉をモデル化するには不十分ということで、こういった忠実な書き起こしをつくっています。

それから、これはうちの特徴ですけれども、先ほど申しましたように、話者によって発音は随分変わりますから、一回普通の音声認識システムで認識して、その結果に基づいて、話者ごとにデータを

衆議院審議コーパスの例

- {えー}それでは少し、今{そのー}最初に大臣からも、{そのー}貯蓄から投資へという流れの中に{ま}資するんじゃないだろうかとかいうような話もありましたけれども、{だけど/だけれども}、{まあ}あなたが言うと本当にうそらしくなる{んで/ので}{ですね、えー}もう少し{ですね、あのー}これは{あー}財務大臣に{えー}お尋ねをしたいんです{が}。
- {ま}その{あの}見通しはどうかということでありますけれども、これについては、{あのー}委員御承知の{その}「改革と展望」の中で{ですね}、我々の今{あのー}予測可能な範囲で{えー}見通せるものについてはかなりはっきりと書かせていただいて{い}るつもりでございます。

区切ります。だれそれさん、だれそれさんと、話者ごとにモデルをチューンしてからもう一回やるので、二回認識するのです。これは当然リアルタイムな処理はできませんが、一回目よりは時間は短くなりますので、実用的にはそんなに問題はなかろうということで、実際には五％ぐらいのゲイン（向上）原因があります。

もっと望ましいのは、総理とか大臣とか、よく発言される方は、それぞれその人ごとにデータを集めてつくることなのですけれども、現実的に可能なのかどうかは、まだ検討しているところであります。

・発音モデル

辞書の方も、話し言葉に特有な発音を入れていかないといけません。

この例での「音声に起因」が、実際の発音の「オンセイニキイン」が「キーン」と長音化することがあります。こういう現象をとらえておくと、「閑静な寺院」という単語についても同様に発音がわかってきます。

これは実際に日本語でどういう現象が多いかということですが、「本当に」が「ほんとに」とか、「それから」が「そえから」、「けれども」は「けども」など、長音化、促音化が多いのですけれども、こういう現象があります。

もう少しおもしろいのは、例えば「用いる・用いれ」のときは「もちーる」と長音化して発音しますが、「用いれ」のときは余り長音化しない。「用いよう」になるとほとんど長音化しないといった現象もとらえております。

・言語モデル

言語モデルとは、言い回し、単語の連鎖がどういうことかということです。

「えー」のようなフィラーも、どこでもかしこでも入り得るわけではなくて、文頭に入りやすいとか、「今」「これは」という後に入りやすいなど、ある程度の特徴がありますので、そういうものもある程度モデル化します。

「このー」の前に「えー」が入ることが多いというのがわかりますと、そのほかにも入りやすいだろうということが予測できます。

これも、今ある百五十時間ぐらいのデータで分析しますと、「えー」が一番多くて、そのほかに「ですね」が多い。置換の例としては、「ている」が「てる」、「という」が「ていう」になることがわかっています。これも全部統計量を求めて、音声認識システムに反映させています。認識するときは「えー」を予測して、後で捨てることをやっています。

こういうモデルをつくるときも、いろいろ検討事項はあるのですが、まずは議会ごとにつくらないといけないということで、東京で話される話題とか語彙と、北海道で話される話題や語彙が随分違いますから、それらを反映する必要があります。

もっといいますと、衆議院でしたら、財務委員会とか法務委員会がありますが、委員会ごとに話題とか語彙も違ってきますので、やってみないとわからないのですけれども、それぞれつくる必要があるのかないのかという検討はしないといけないと思います。

また、きょうの会議ではこういう質問がなされるということが事前にわかっていましたら、質問書、答弁書から予測もできます。ＮＨＫのニュースは、当日のアナウンサーの原稿を機械に入れて

実際にやっていますけれども、それに近いことをやるとよくなる。よくなるとは思うのですけれども、こういうことをやると手間がふえますので、どのくらいよくなって、どのくらい手間がふえるかというトレードオフも考えなければいけません。

こういったことを反映して開発しておりまして、現時点では衆議院の予算委員会の総括では、八〇％ぐらいの認識精度になっているということで、現在これの改善を目指して研究開発を行っております。

私は、予算委員会がほかと比べて難しいかどうかはよくわかっておりませんけれども、本会議よりは予算委員会の方が難しいのは自明でして、ほかの外務委員会とか法務委員会に比べてどうかというところは、まだ検討はしておりませんが、そんなに大きく違いはないだろうと考えております。

・デモンストレーション

ここで、どのぐらいのものかというのを体感していただきたいと思います。七〇％、八〇％、九〇％という順番にお見せします。最初が七〇％の例です。

この議論は、ちょっと精査しなければならぬと思いますが、例の米朝の合意ができましたね、一九九三年か一九九四年かな。あの時点から、アメリカは北に優しくなりましたね。韓国も、北に優しくなりましたね。日本もコメなんか支援したりして、優しくなりましたね。結局あの時点で、北朝鮮の核は大体おさまったなと、みんな安堵の胸をなでおろして、その後は、それ以上激しくすると、また彼らが機嫌を損なうという意味で、結構太陽政策みたいなものが始まったのは、そのあたりからだと思うんです。

しかし、そういうことをやりながらつき合ってきたら、ふたをあけたら核という問題が出てくる。そういう意味で、これは交渉次第だと思いますけれども、特に経済協力等については、あるいは人道主義的なものについては、変に施すみたいな気持ちにならずに、核は核として検証して、そのことが本当に北朝鮮の核を放棄するような方向に動くのかという、そこだけは本当に精査した上で心を、気配りをしてもらいたいなというのが、私の心からの願いでございます。

これはえらくタイムリーな話題なんですが、実際には二年前の委員会のデータです。七〇％ではなかなか使いものにならないのではないかと思われる方も多いのではないかと思います。

次に八〇％の例です。

○答弁　ＥＴＦに関しましては、これは元本保証のない商品でございます。そのことは事実として重要な点であるかと思います。

不適切な発言に不適切な部分があったということを踏まえ、以後、注意をいたしたいと思います。

○議長　海江田万里君。

○海江田万里君　**これは大いに反省をしていただかなければいけないわけですが、そういう形で、一日も早く削除した方がいいですよ、本当に、これはね。それはやっていただけるわけですね。それはぜひ、本当に一日も早くお願いをしたいと思います。**

それでは今、最初に竹中大臣からも貯蓄から投資へという流れの中に資するんじゃないだろうかとかいうような話もありましたけれども、だけどあなたが言うと、本当にうそらしくなるので、もう少し、これは塩川財務大臣にお尋ねをしたいんですが、もう片一方で、個人国債、個人向けの国債が大変売れたということですが、その個人向けの国債が売れたということが、この貯蓄から投資へという方向に役立っているのかどうなのかということで、ご意見を伺いをしたいわけですが、これについてはどうでしょうか。

できているところとできていないところが随分はっきりしていると思うのですけれども、やっぱり早口のところが難しいという感じはするかと思います。

最後に、九〇％の例です。

○赤嶺君　**査察を成功させるという立場に、外務大臣、立っておられるのか、あるいは査察について、今どんなふうに考えているのか極めて不明なんですが、査察を成功させようという立場ではないんですか。**

○議長　川口外務大臣。

○川口外務大臣　**私は、ぜひ査察を成功させたいと思っています。そして、その成功させたいという国際社会の気持ちをイラクが理解をして、成功させるための能動的な協力をする必要がある。それがなければ査察は成功しない、そういうことだと思います。**

こんな感じで、みんな川口さんのようにしゃべってくれると非常にありがたいのですが、この九〇％の例ですと結構使いものになるかなという印象をお持ちいただけたかと思います。

主観的に申しまして、九〇％ぐらいだったらいいかなと思いましたけれども、七〇％ぐらいではちょっとつらいのではないか。八〇％ぐらいだと、微妙だなと思われた方が多いのではないかと思うのですが、先ほど申しましたように平均で八割ぐらいで、目標としては九割に近づけるべく努力をしているという状況であります。

認識は書き起こすだけではありませんで、幾つか関連するトピックとしまして、雑音とか残響対策、特に反響、エコーが入るところの検証も必要かもしれません。音響機材をよくするかどうかで、認識率が随分変わってくるとも聞いております。それから、話者の特定をする。実際に会議録を作成される際には、発言者も入れないといけませんので、話者のラベルが自動で入るといい。

先ほどの例では、議長の指名の発話のチャンネルは落としていましたけれども、それを音声認識できるかもしれませんし、話し手の特徴をとらえてラベルをつけるということも、技術的にはそんなに難しいことではありません。

それから、「えー」とか「あー」といったフィラーが随分多いので、それも自動的に除去することも必要かと思います。

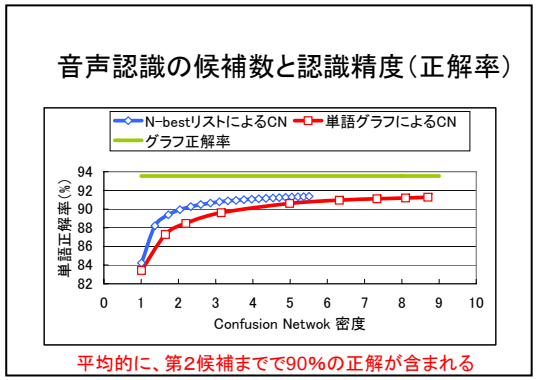
・音声認識結果の情報

以上が音声認識の技術的な話ですけれども、最後にこれを会議録作成にどのように適用していくかにつきまして、幾つか衆議院の方とも何回か議論を持たせていただいたことも踏まえまして、現在考えていますことを御紹介いたします。

まず音声認識システムの課題としましては、今ビデオで見ていただいたような認識結果の第一候補だけではございませんで、第二候補、第三候補も出ております。

それから、認識には一応信頼度がある。この辺は自信を持って出しているけれども、この辺は怪しいということもある程度ついております。

一番いいのは、単語ごとに対応する音声の区間がわかっておりますから、認識が怪しいところ、あるいは認識が間違っていると判断されたところのみをピンポイントで再生することができます。あるいは、そこだけゆっくり再生することもできます。



平均的に、第2候補までで90%の正解が含まれる

このグラフは、複数候補を出したときにどのぐらいの認識率になるかというもの。例えば青い方を見ていただきますと、認識率の数え方が若干違うのですけれども、第一候補で八二%ぐらいのところを第五候補ぐらいまでとりますと九〇%をちょっと超える。第三候補ぐらいでも九〇%を超えることがおわかりいただけるかと思います。

・音声認識結果の利用法

こういう情報を踏まえまして、どのように使えるかということについて、幾つか紹介いたします。

まず一番の方法としまして、ワープロに先立って人手修正を介するテキスト化を行ってしまう。これは北海道議会さんを含めまして、幾つか入れられているところはこういうやり方でやっておられると思います。とにかく最初に完全なテキストをつくってから、後で会議録としての体裁を整えていく。このメリットとしては、ワープロと音声認識修正ソフトを分けられるということがあります。テキスト化する処理と、後で会議録をワープロで作成する段階と、二段階のフェーズが入ります。

もう少し具体的にいうと二通りありまして、北海道議会さんとかで使われていますアドバンストメディアの商品では、第一候補のみを表示して、だめならタイプで修正します。

それに対して、京都大学で試しにつくってみたものでは、せっかく第二候補まであれば九〇%まで行くのですから、それを表示して選択できるようにすればいいのではないかということにしました。ワープロの仮名漢字変換で第一候補がだめで、何回かキーをたたけば第二、第三候補で出てくるのと同じようなインターフェースになっています。

アドバンストメディアさんの「リライター」という商品の画面をいただいてきます。特に宣伝ではありませんで、参考のために見ていただければと思います。

音声波形がございまして、今、編集しているところの認識結果が出てきます。例えばここがおかしいというところをクリックすると、そこの音声が聞けるようになっていまして、画面で音声を聞きながら、間違ったところを逐一直してい

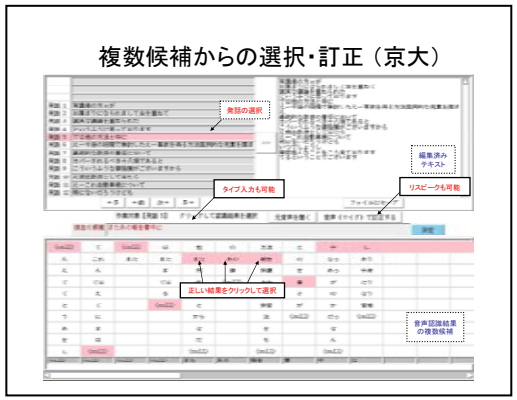
きます。

この原稿をお出ししたのが一カ月ぐらい前だったのですけれども、十月になってから、ＮＥＣさんの方が同じような、「同じような」と言うと怒られますけれども、議会向けのソリューションシステムを発表されたので、それをちょっといただいてきました。

見た目は、似た感じですけれども、テキストの編集機能がかなり充実しているという違いがあります。例えば発言者の自動推定もやっていますし、ラベルも入れられるようになっている。あと、会議録で必要な「（拍手）」、「（起立）」といったコメントも、この画面で入れられます。

また、これは商品ではないのですが、うちで試しにつくってみたインターフェースです。さっきと同じで、最初の変換が間違っていたら第二候補、第三候補を出すやり方です。それを一々クリックしていたら手間がかかりますから、最初から複数個を出して、そこから選んでいけるようになっています。

さっきのグラフで申しましたように、全部合わせると九割ぐらい正解が含まれているということで、九割というのは、さっきの川口さんの例でありましたように、結構イケているという意味で、こういう方法もあるのかなと考えています。



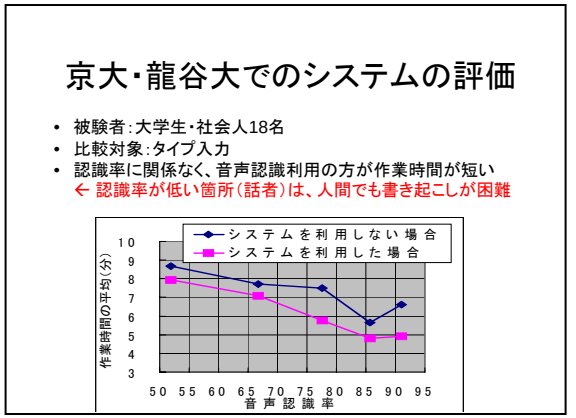
実際には、何人かの方とお話ししますと、こういうやり方は余りよろしくなくて、先ほどのやり方の方が効率がいいと聞いております。タイプ歴にもよるのですけれども、速記者のようにタイプがなれている方でしたら、キーボードだけで全部仕事が済む方が速いということで、私もそれはそうかなと思います。

今お見せしたのが、世の中にある商品というかシステムですけれども、もっと別のやり方の可能性として、例えば音声認識結果をいきなり出さないで、人間がタイプをしていくと、音声認識結果を引っ張ってきて、合っている候補を逐次出すやり方もあるかと思います。

携帯電話で入力をするときに、「おは」と入れれば「おはよう」と出るとか、予測入力に近い考え方です。ワープロソフトなら一回で済みますから、作業手順をふやすことなく音声認識結果もそれなりに効果的に利用できるのではないかということですが、ソフトをつくるのが面倒くさい、あるいはワープロソフトのバージョンが変わったら使えなくなるのではないかといった安定性、将来性に関して懸念があります。

もう一つの方法としては、全部音声認識結果を使うのではなくて、先ほどの海江田さんの例でごらんいただきましたように、できているところはほとんど合

っていますし、特に早口になったところはほとんどできませんので、できているところだけ音声認識結果を出して、できていないところはタイプすると、ストレスも少なくなるのではないかと考えられます。個人の好みに応じて閾値を変えることも考えられるかと思います。



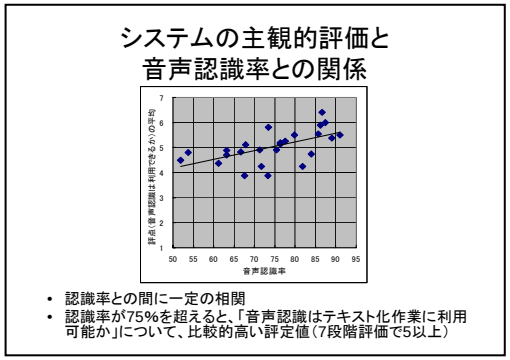
先ほど申しました今できているもので、実際に京大と龍谷大学の学生さんに試しにやってもらった例であります。

大学生が多いので、皆さんのようにプロの速記者と、タイプ入力のスキルが随分違うと思うのですけれども、一応参考にはなるかと思います。

課題としては、先ほどお見せした衆議院の予算委員会の音声認識結果をもとに、会議録を作成してもらいました。横軸が認識率で、五〇％から九五％までランダムに選んでいます。

赤い方は音声認識を使った場合で、黒い方は一から聞いた場合。聞くところはさっきと同じような画面を使っていますので、違いは音声認識結果からスタートしているかいないかだけですが、これは僕らにとってはグッドニュースで、認識率には関係なく音声認識結果を使った方が、作業時間は短くなっています。

興味深いのは、認識率が高いと、全部人手でやっても、機械でやっても速くなりますし、認識率が低い人は、機械でやっても、全部人手でやっても時間はかかる。認識率が低い箇所、あるいは低い人は、人間が書き起こしても難しいということが言えるかと思います。



これは以前、衆議院の方とお話ししたことがあるのですけれども、議員さんごとに認識率のグラフを出して、それをお見せしたところ、「この先生は難しいんですよね」ということをおっしゃっていました。機械がやっても難しいところは人間がやっても難しいということが言えます。

このグラフは、認識率と被験者の方の主観的な評価の関係をプロットしたものです。一応七段階評価で、使い勝手がいいかどうかを評定してもらいました。四が真ん中なので、五以上がいいと思われるところで、五を超えるところは大体七五％ぐらいです。

これで言えるのは、使っている側としては、認識率が高い方が使い勝手がいい。さっきのグラフとは違う結論ですけれども、認識率の高い方が、このシステムの使い勝手がいいと感じられる傾向にありまして、その閾値が七五％ぐらいのところにあります。

一応タイピングスキルとも関係がありますので、それとの相関を見たところ、パソコン使用歴の浅い人の方が、このシステムを使い勝手がいいという傾向にありました。そういう意味で、プロの方が使うとどうなるかというのは、今後評価検証していかないといけないところになります。

・まとめ

話は以上です。これは私の願望に近いところもあるのですが、品質面で、例えば単語とか固有名詞を音声認識の辞書に登録しておきますと、特に固有名詞の揺れがなくなるという副次的な効果もあります。

それから、さっきの繰り返しですが、熟練した人が集中した場面において作業時間が短くなるかというのは今後評価検証していく必要があるのですけれども、少なくともタイプし続ける状況はなくなりますから、長時間作業の負荷は少なくなるかと思います。

具体的に言いますと、今、例えば衆議院では五分ぐらいのデータが限界ですけれども、こういうシステムを使うと、もうちょっと長くやってもそんなに疲れないのではないかと期待しています。

もう少し明らかなのは、このシステムを使う訓練が今の速記を学習するのに比べて少なくて済むであろうということです。

以上をまとめますと、話し言葉の音声認識は一般に容易ではありませんし、それができたかできないかと言われると、私は「できない」と答えざるを得ないと思いますが、議会のような特定の分野のオーダーメードにすると、それなりにできつつあるということで、特に議会の音声認識は、実は比較的その中でも容易な分類になります。それは、膨大な会議録などの学習データが存在するからです。

実際に衆議院の会議録というのは、年間に千八百万～二千万単語ぐらいの量がありまして、それは数えると、大体新聞記事一年分と同じです。衆議院の速記者の方がつくられているテキスト、あるいはその議員さんが話されている内容は、新聞と匹敵するぐらいの量があるのです。

もう一つは、収録条件が比較的いいということ。マイクも音響設備も、ちまたのものに比べれば、ちゃんとセッティングされています。話者も、基本的には話しなれた方が丁寧に発声することで、八〇～八五％ぐらいの認識精度は可能と考えられます。しかしながら、それが使いものになるかというところに関しましては、まだ幾つかの事例はありますけれども、今後一層の検討が必要になります。

どうも長らくおつき合いいただきまして、ありがとうございました。以上で私の話は終わらせていただきます。（拍手）