…平成22年度助成研究より…

高認識率・低騒音を達成する音声認識用マスク型マイクの開発と応用領域研究

吉備国際大学　社会学部
ビジネスコミュニケーション学科
教授　佐藤　匡

1．はじめに

障害者自立支援法第77条により、聴覚障がい者への情報保障が自治体の義務となっている。情報保障とは、あえて一言で言えば、その場の会話や発言を手話や文字などの視覚情報として聴覚障がい者に伝えることである。

本研究は、音声認識をこの情報保障に利用するための研究と、その応用の可能性について検討するものである。

2．要約筆記による情報保障

補聴器は万全ではなく、また聴覚障がい者全員が手話を使うわけではない。今日、手話と並んで多く利用されている情報保障の手段は要約筆記である。文字通り紙に筆記したものを読んでもらうわけだが、情報保障にはリアルタイム性が要求される。人が話す速さ（約300字／分）とそれを文字にして書くのには速度差があるため、要約して文章化せざるを得ない。だから「要約筆記」である。要約筆記のサービスを提供する人を要約筆記奉仕員とかノートテイカーと呼んでいる。

要約筆記にはおもに２つの方法がある。ひとつが手書きの要約筆記でもうひとつがパソコン要約筆記である。手書きで提示できる情報量はベテランで60字／分程度であり、パソコンのタイピングだと150字／分程度である。パソコン要約筆記では手書きの２倍以上の情報量を提供できるが、設営などに制約がある。どちらにしても数十分も連続して要約筆記するのは集中力のいる作業であり、ノートテイカーには訓練や慣れが必要である。

3．音声認識による要約筆記の課題

多くの大学が聴覚障がいの学生向けに講義の情報保障を行っている。講義を要約筆記する場合、講義内容や専門用語を知っている大学生が要約筆記をするのが望ましい。しかし、タイピングの育成に時間がかかる学生ノートテイカーは４年で卒業していく。このノートテイカー不足の慢性的な問題を解決するために、他大学の学生ノートテイカーにインターネット経由で音声を送り遠隔入力して文字化してもらう試みや、パソコン要約筆記に音声認識を併用させて、タイピングの負荷を下げる試みが行われてきた。

遠隔入力については、各教室のファイヤウォール越え設定の煩雑さや遠隔入力の体制などの問題があり実用化が進まない。一方、音声認識は手軽な解決手段であるため工夫次第では活用の余地がありそうである。この原稿の執筆時点において、個人でも購入が可能で簡単に使え、最も認識率が高い音声認識ソフトはAmiVoiceである。

パソコン要約筆記にこのソフトを併用する場

合、リスピーク方式を取るのが賢明である。「リスピーク」とは、教員の音声を直接認識させるのではなく、特定話者としてあらかじめソフトに声を登録しておいたノートテイカーが、教員の話す内容を要約しながら復唱して文字化していく方法である。リスピークにする利点は３つあって、ひとつは特定話者として最初から認識率を上げておくことができること、ふたつ目は、文法的におかしな文章をノートテイカーが復唱時点で修正したり、ことばのけば（無機語）を除いたりすることができることである。「○○とか△△とか、あと□□とか、そうえいば××も……それで次のページを見ると……」といったような文法的におかしい話ことばは音声認識ソフトには大敵であり、また「あのー」「あー」「えーっとー」といった無機語ももちろん大敵である。慣れていない人が、アナウンサーのようにことばを選びながら無機語を極力使わないように話すのはかなり難しいことである。息継ぎのときには「間」が必要であり、そのときに無意識に無機語を使ってしまうことが多い。息継ぎをアナウンサーのように短くしようとすると、次第に息苦しくなってくる。話しことばは揮発することばであるため、言い直しや無機語などが大量に含まれていても聞き手は無意識にその部分を除外して聞いている。ところが音声認識ソフトにはこの「除外する」ということが非常に難しい。

３つ目の利点は、ノートテイカーがことばを噛んでしまったり発音が不明瞭だったりして誤認識させてしまった場合でも、ノートテイカー自身が文章を削除するなり言い改めたりといった修正作業をすることができることである。教員が誤認識を自分で修正しながら講義を進めるのは不可能であろう。

リスピーク方式はいいことばかりのようであるが、当然問題点もある。ひとつは教員や周囲の学生の声、騒音による誤認識である。教員の声をノートテイカーのマイクが拾って誤認識を起こすし、教室内外の騒音や周囲の学生の声なども誤認識の遠因となる。もうひとつはリスピーク音声の問題である。講義をノートテイカーが復唱するので、常にボソボソと私語をしているような状況となる。教員も周りの学生も講義に集中できなくなる危険性がある。３つ目は、ノートテイカーにリスピークや誤認識修正の訓練が必要だということである。タイピングの負荷は減るが、音声認識との併用に慣れておく必要がある。

４．マスク型マイクによるソリューション

上記のような制約を踏まえて、マスク型のマイクを使ったリスピーク方式による音声認識を試みた。図１のように試作品を実際の講義に試用している。形状は市販の防塵マスクからヒントを得たもので、性能と小型化を優先させている。

図１　2008年度の試作品と、講義での試用例

図２　2010年度の試作品

今年で試作は９タイプ目となる（図２）。当初の課題であった１：マイク部構造の最適化、２：視野を妨げない形状、３：マスク内への侵入騒音の抑制（マスク材質検討）、４：マスク外への音漏れの抑制、５：呼気の湿気対策については解決しており、今回新たに６：軽量化、７：留め具の簡素化、８：弁の低負荷化と脱着機構について検討した。画像ではわかりにくいが、呼吸と防音のための弁部位を大幅に改良している。

AmiVoiceに付属しているマイクとこのマスク型マイクとでの認識率を比較し分散分析した

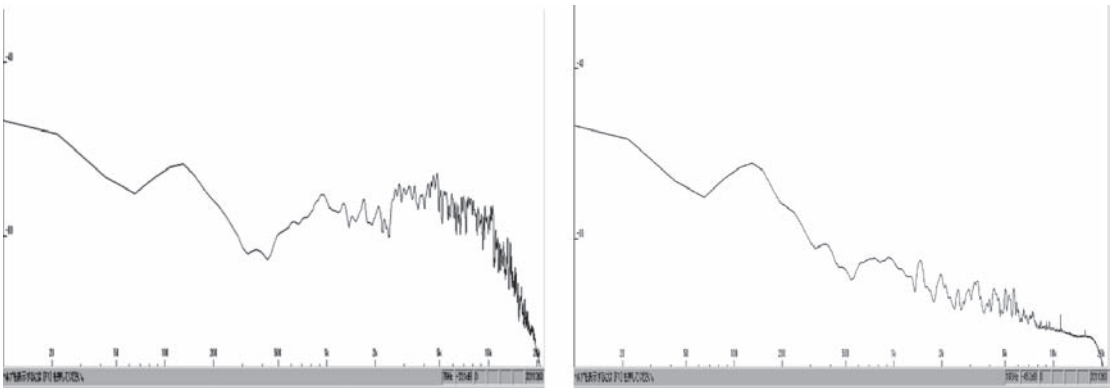

図３　マスク外部の音（左）とマスク内部への音の混入（右）

ところ、両者の認識率には差がないという結果を得た。教室内での使用についても、教員の声にかき消されたこともあって周囲の学生はリスピーク音声に気付かなかったということであった。マスク外部から内部への音の遮音性能については約－20dB@2kHzの減衰結果が得られた（図３）。

５．おわりに

現状の音声認識技術は、文法的に適切な話ことばと雑音のない環境においては高い認識率が達成されている。しかし、話の途中での言い直しや無機語、音声認識辞書にない単語、方言、主語述語が不整合を起こしているような文章・話しことばには弱い。

誤認識は、一旦文字化されてしまうと厳しい評価の目にさらされるが、状況によっては100％の認識率でなくても構わないはずである。誤認識をどこまで許容するか、許容できるのか、誤認識でもいい部分はどこか、そういった議論も重要であろう。

本研究の今後であるが、短時間での準備と撤収、原因不明（ちょっとしたマイクのずれなど）の認識率の低下を起こさないこと、などについての検討も必要である。引き続き総合的な性能を向上させていきたいと考えている。

本稿を執筆した2010年11月、偶然、情報処理学会誌「情報処理」No.11に音声認識技術の最新の状況が詳しく紹介された。関心のある方はぜひそちらをご覧いただきたい。