

J-007

音声認識を用いた All for one 型多言語会議支援システム SAKIN の開発

SAKIN: All-for-one Type Multilingual Conference Support System Using Voice Recognition

井出 美奈[†]
Mina Ide

重信 智宏[†]
Tomohiro Shigenobu

吉野 孝[†]
Takashi Yoshino

1. はじめに

近年、在日外国人の数は増加してきており、多言語間でのコミュニケーションの機会が増えてきている。また、平成19年5月における留学生の数は、約12万人に上っている[§]。講義や会議などの場面で、言語の違いにより外国人の内容の理解には限界がある。そのため多言語間の会議の場を支援するシステムが必要である。

日本国内における会議の参加者の大部分は日本人であると考えられる。日本人は参加者であるが、外国人の会議の内容理解のための支援者にもなり得ると考えられる。本研究の「All for one」とは、参加者の多くの日本人が協力して外国人を助けることを指し、本研究の目的としている。

これまで利用者間において会議の情報を共有するシステムとして、Langrid Blackboard[1]、EGITool[2]などが提案されている。我々は、Langrid Blackboard を使用して多言語会議の要件抽出実験を行った [3]。その結果、支援者の作業の競合や、発表内容に対して支援作業に遅延が発生するといった問題が挙げられた。

そこで、支援者の入力負担を軽減するために、音声認識を使った多言語会議支援システム SAKIN を開発した。しかし現在、音声認識の精度は十分ではなく、音声認識結果を支援に使うには、音声認識結果の修正や使用方法の工夫が必要である [4]。

本稿では、音声認識で得られた情報から、有用な情報を抽出して支援に使うためのユーザインタフェースを提案する。提案方法を実装した SAKIN を多言語会議に適用し、ユーザの評価および結果の考察について報告する。

2. 多言語会議での要件

我々が行った Langrid Blackboard を用いた実験 [3] より、以下のような要件が挙げられた。

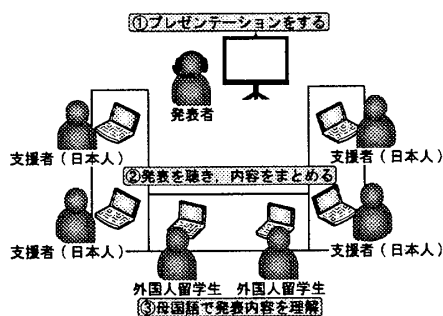


図 1: SAKIN を使用した多言語会議

[†]和歌山大学システム工学部

[‡]独立行政法人情報通信研究機構

[§]<http://www.jasso.go.jp/statistics/>

- (1) 発表内容に対して支援作業の遅延が発生しないよう入力時間の短縮を支援する
- (2) ユーザ同士の作業の競合を防ぐため、他のユーザの入力中の行動を共有する
これらの問題を解決するために、SAKIN の開発を行った。

3. 多言語会議支援システム SAKIN

(1) SAKIN の基本的機能

SAKIN は図 1 のように多言語会議で使用する。プレゼンテーションを行う発表者が発言した内容を音声認識し、単語ラベルが自動生成され、図 2 の情報共有フォームに表示される。情報共有フォームに表示された単語ラベルや、グルーピングのためのオブジェクトや矢印などのオブジェクトは、インターネットを介して共有される。それらのオブジェクトに含まれる内容は言語グリッドを介して、リアルタイムで外国人聴講者の各母国語に翻訳される。各支援者が単語ラベルを修正したり、配置して整理したりすることにより、要約された議事録を作成していく。入力は折り返し翻訳機能を備えた入力フォームから行う。

(2) 色によるユーザ識別機能

各ユーザに赤・青・緑等の色を図 2(3) の様に自動で割り当てる。ユーザが単語やグループの移動、文字の修正等の操作を行ったとき、図 2(2) の様にオブジェクトの色がユーザに割り当てられた色に自動的に変わる。これにより、他のユーザの作業が視覚的にわかるようにし、ユーザ同士の作業の競合が起こらないようにする。

(2) 単語寿命機能

音声認識結果をそのまま支援に使うと、発表者の発言内容とは異なる可能性がある [4]。一般的には、高精度な音声認識のために、認識内容に合わせた専用の音声認識用辞書を作ることが多い。しかし、専用の音声認識用辞書は、会議の内容に依存するため、予め準備すること

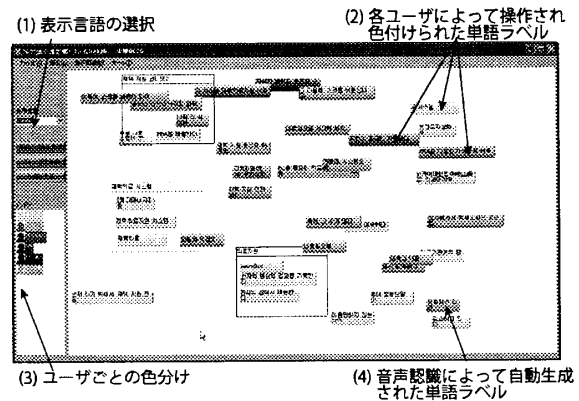


図 2: SAKIN 情報共有フォーム

表 1: 日本人支援者のアンケート結果

アンケート項目	A	B	C	平均
(1) プレゼンテーションが速すぎて入力が間に合わなかった	2	3	4	3.0
(2) 自分が書こうと思っていた内容を他人に書かれてしまった	3	2	4	3.0
(3) 自分が書いた文を勝手に消されたり、勝手に編集されたりして嫌だった	4	3	2	3.0
(4) 他の人の行動が予測できた	1	2	2	1.7
(5) 音声認識により自動的に単語が入力されることで、入力がしやすくなった	5	2	3	3.3
(6) 音声認識により自動的に単語が入力されることで、画面が見にくくなった	2	5	4	3.7

評価尺度 1:強く同意しない 2:同意しない 3:どちらともいえない 4:同意する 5:強く同意する

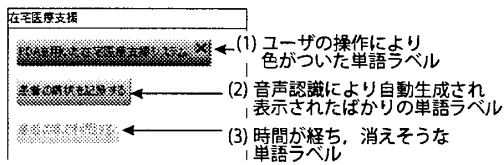


図 3: SAKIN 単語寿命機能

は困難である。

そのため、一時的に音声認識結果の全てを、単語ラベルとして自動生成して、その中からユーザが必要だと考えた単語ラベルのみを利用する仕組みを考えた。図 3(3)のようにユーザに操作されなかった単語は自動的に色が薄くなっていき、一定時間経つと消えるようにした。また、消したくない単語はユーザがマウスで触れることにより元に戻り、消えるのを防ぐことができる。これにより様々な情報の中からユーザが使いたい単語ラベルだけを残すことができる。

また、マウスで触れて元に戻る際に、単語ラベルに各支援者に割り当てられた色の枠が付く。これにより、他のユーザがどの単語ラベルに注目しているかが視覚的にわかる。

4. 音声認識を用いた多言語会議支援実験

4.1 実験方法

音声認識を用いることで多言語会議での要件を満たすかを検証するために、多言語会議支援実験を実施した。実験の被験者はプレゼンテーションを行う発表者が1名、支援者は3名、留学生は2名である。実験時間は17分であった。留学生の母国語は韓国語で、閲覧のみの利用とした。留学生は、全員日本語の日常会話ができる。実験終了時にアンケートを実施した。

4.2 実験結果

実施したアンケートは5段階評価と自由記述式の両方を用いた。支援者への5段階評価のアンケートの結果を表1に示す。

1) 入力時間の短縮について

表1(5)の回答よりばらつきはあるが、音声認識で単語が自動生成されることで入力しやすくなっていると考えられる。日本人支援者の自由記述の回答によると「発表内容と少し違うだけなので、新規作成ボタンを押さずにすむので楽だった」といった意見が得られた。

2) ユーザ同士の作業の競合について

「今操作している周辺の他の人の作業は、色によりわかった」といった意見が得られた。しかし、表1(2), (3), (4)の回答より、ユーザが他のユーザの作業をあまり把握できていない。また、音声認識単語がリアルタイムに自動生成されたことにより、出てきた単語ラベルに対す

る操作が集中していた。「操作をしようとした単語ラベルが今話している内容のため、他の人に書かれてしまった」といった意見が得られた。

また「今操作している周辺はわかるが、次の操作がわかりにくい」といった意見が得られた。

3) 画面上のラベルの散乱について

「それほど量も多くないので、自動生成ラベルは邪魔というほどではなかった」といった意見が得られた。しかし一方で、表1(6)の回答より、音声認識で自動生成した単語ラベルが多くなりすぎ、操作しづらいという問題があると考えられる。自由記述の回答からも「自動生成ラベルが散乱し、入力よりも整理に力を注いだ」といった意見が得られた。

4) 外国人留学生の回答について

「画面が見やすかった」といった意見がある一方で、外国人留学生2名のどちらからも「急に単語が出てきて、どこを見ていいのかわからず混乱した」といった意見が得られた。これは、音声認識で自動生成した単語ラベルが散乱したため、発生した問題であると考えられる。

5. おわりに

本稿では、音声認識を利用して多言語会議を支援するシステム SAKIN を開発し、多言語会議支援実験を行った。実験の結果、ユーザの入力の負担を軽減する可能性があることがわかった。しかし、今回実装したユーザインタフェースだけでは競合を防ぐことや、音声認識精度を補う役割は不十分であったといえる。

今後の課題として、支援作業の競合を防ぐために音声認識結果の表示方法を検討する予定である。また、音声認識の精度の不十分さを補う新しい方法として、音声認識結果から Google 大規模日本語データを用いて、会議や講義の内容に合う専用の音声認識用辞書の作成を検討する予定である。

参考文献

- [1] 重信 智宏, 藤原 義功, 村上 陽平, 吉野 孝, 石田 亨: ドメイン指向翻訳を備えた言語グリッドコミュニケーションツール, 情報処理学会第 69 回全国大会講演論文集, 3G-1, pp.67-68(2007).
- [2] 江木啓朗, 石橋啓一郎, 重野寛, 村井純, 岡田謙一: 共同記録作成を基にした対面議論への参加支援環境の構築, 情報処理学会論文誌, Vol.45, pp202-211(2004).
- [3] 井出美奈, 重信智宏, 吉野 孝: 言語グリッドを用いた多言語会議支援システムの要件, FIT2007 情報科学技術フォーラム, 第3分冊, pp.483-484 (2007).
- [4] 田窪行則, 前川喜久雄, 窪蘭晴夫, 本多清志, 白井克彦, 中川聖一: 言語の科学 2 音声, 岩波書店 (1998).