

発話者の方向と反応を表示する音声認識アプリが示す情報保障の可能性

大杉 豊
筑波技術大学
osugi@a.tsukuba-tech.ac.jp

概要

きこえる人複数による会話や会議に参加する際、手話言語通訳者の配置によって参加者の発言（音声）内容を知ることができる。しかし、発言者が誰であるか、また発言者以外の参加者の反応（音声）などの情報をリアルタイムで得るのが困難という課題がある。通訳者もこれらの情報を瞬時に伝えるに限界がある。本発表では、音声認識アプリの新製品「VUEVO」のインターフェイスがいかにして上記の課題を解決するかの事例を紹介し、情報保障の新しい可能性を考える。

1. 問題の所在

筆者は、本学できこえる教職員や研究者数名とテーブルを囲んで会議や議論をする場面では、手話言語通訳者を配置をして意思疎通を図るのが常である。以前に音声認識アプリの併用を試みたことが数回ある。しかし、きこえる人の発言行為を発言内容はもちろんのこと、それ以外の要素も含めて、会話全体の流れを把握することが最も重要であるとの考えにより、音声認識アプリの併用は断念した。具体的には、発話の重なりや相槌を含めて、「誰が」「どんな内容を」「どのタイミングで」発声したのかの情報を音声認識アプリで生成される文字画面からは得られないということである。これらの情報は手話言語通訳を通して得ているが、その情報の質が通訳者によって変わってくるために、発言内容を把握できても、会話全体の流れを把握する作業が安定しない問題がある。

2. 新製品「VUEVO（ビューボ）」の特徴（ピクシーダストテクノロジー株式会社のウェブサイトより）

複数人での会話や会議の際、発話内容を360°全方向から集音発話した内容をリアルタイムに発話者の方向に表示

独自技術で開発した精度の高いワイヤレスマイク

ピクシーダストテクノロジーの独自技術で、設計・開発したワイヤレスマイクです。限られた内部スペースに8つの高性能マイクを内蔵し、360°全方向から音声を集音しながら発話者の方向を特定。精度の高い音声認識で正確なテキスト変換と方向表示を実現します。マイクは軽量・コンパクトで携帯しやすく、ワイヤレスで使えます。



360°高精度な集音

8つの高性能マイクを搭載し、集音特性を高める設計。環境ノイズの影響を受けにくく、精度の高い音声情報の検出を可能にしています。

音源特定アルゴリズム

雑音や室内の反響に強い音源特定アルゴリズムを開発。高精度かつリアルタイムな発話方向特定と音声テキスト表示を実現しました。

信号分離アルゴリズム

複数音源の分離と検知を可能にする、空間的信号分離アルゴリズムを開発。最大3方向からくる音声情報を同時に認識しテキスト表示が可能です。

複数人の会話内容を直感的に理解しやすいインターフェース

360°ビューの直感的なユーザーインターフェースにより、会話や会議の進行中でも発話の内容と共に発話者の方向をわかりやすく表示します。最大8方向の表示が可能で、視覚的に見分けが付きやすい専用カラーを各方向に配色しています。テキストのフォントサイズも設定からかんたんに変更可能で、ユーザーの好みや共用の画面でビューを投影し使用する際にも最適なサイズに調整できます。



3. 本学におけるトライアルで得られた活用例

2023年6月22日から8月31日まで本学にて「VUEVO」のトライアルを実施した。筆者がきこえる教職員や研究者数名とテーブルを囲んで会議や議論をする場面での試用は14件中4件であった。この4件における筆者の「VUEVO」活用例について述べる。筆者がきこえる教職員4名とテーブルを囲んで会議をする形で、筆者の真正面に手話言語通訳者が位置する（右図を参照）。テーブル上、中央部分に「VUEVO」の集音マイクを置き、その手前にiPadを設置して、「VUEVO」アプリの360度ビューインターフェイスが画面に出るようにした。その結果、筆者の席から見ると手話言語通訳者が手腕を動かす範囲のちょうど下に重ねるように iPad の画面を見られる形となる。

筆者は手話言語通訳者の胸上あたりに視線を置くという基本を変えないが、発声を瞬時に色で表示する「VUEVO」の360度ビュー画面が視野の下部に入っているため、「誰が」「どんな内容を」「どのタイミングで」発声したかという情報を難なく得ることができる。参加者それぞれの発声情報が異なる色と角度で画面に示されるため、発声が重なるなどのタイミングを自然に把握できる点、そして「うんうん」「そっか」など相槌を打つような内容の文字も出される点を含めて、手話言語通訳者に向けている視線を動かさずにこれらの情報をほぼ得られることは筆者にとって初めての経験であった。また、必要に応じて画面に出されているテキストを確認するときも視線を最小限に動かすだけで「誰が」の情報を含めて確認できる点も優れている。

4. 今後の可能性

「VUEVO」は、筆者がきこえる教職員や研究者数名とテーブルを囲んで会議や議論をする場面で、手話言語通訳配置と併用することで、「誰が」「どんな内容を」「どのタイミングで」発声したのかの情報を得るに大変有効なシステムであることがわかった。何よりも重要なことは、ろう者が会議や議論などに完全参加を果たすために必要な情報保障のあり方を、当事者も主体的に開発者と連携して追求していくことである。今後も本学で「VUEVO」の活用事例を重ねていくことで、活用できる場面とそれぞれに必要な機器の有無を整理していくとともに、学生のキャリア教育への導入の可能性を探っていきたい。また、手話言語通訳者の視点では、「VUEVO」の併用により、「誰が」と「どのタイミングで」の情報や参加者全ての発声情報を通訳者だけが伝える必要がなくなることで、負担が軽減されることが予想されるだろう。



「VUEVO」に関する詳細は

ピクシーダストテクノロジー株式会社 VUEVO（ビューボ）

