

音声認識を使った文字による情報保証の現在と未来

Present and future of the information assurance by the text using speech recognition

福岡 寿和

Toshikazu FUKUOKA[†]

[†] 株式会社富士通ソーシャルサイエンスラボラトリ [†] FUJITSU SOCIAL SCIENCE LABORATORY LIMITED

E-mail: [†] ssl-info-livetalk@cs.jp.fujitsu.com

1. はじめに

富士通は、1990年代後半からICTを使ったユニバーサルデザインの推進に力を入れており、障がい者のユーザビリティを考慮したスマートフォンなどの製品を市場に提供している。このような視点で投入した製品は、アクセシビリティが向上しており、障がい者にとって使いやすいだけでなく、健常者にとっても使いやすいという特徴がある。

本稿では、リアルタイム音声認識による複数発話者による会話をタイムラグなく表示する「Fujitsu Software LiveTalk（以下、LiveTalk）」の開発、実証実験、お客様ご評価を通じてみてきた情報保証の現状とあるべき未来の姿について述べる。

2. 聴覚障がい者への情報保証環境

LiveTalkの開発がスタートした2013年当時、企業内における聴覚障がい者への情報保証環境としては、次のようなものが一般的であった。

- ① 外部要約筆記者による発話内容の要約筆記
- ② 外部手話通訳者による発話内容の手話通訳
- ③ 隣席社員によるパソコンを使って要約筆記
- ④ 隣席社員による要点メモでの筆記通訳

筆者は聴覚に障がいを持った同僚と一緒に仕事をすることがあり、日々のコミュニケーションや会議での情報保証として、主に電子メモパッドを使った筆記通訳を用いていたが、次のような課題があった。

- ① 要点メモが会議のスピードに追いつかない
- ② 筆記通訳者が会議の内容を把握しづらい
- ③ 要点メモでの指示に対して不明部分があるかどうか把握しづらい

また、聴覚に障がいのある社員からも「会話にリアルタイムについていけない」、「作業を進めていく過程で要約にはない部分について把握している必要があった」、「わからないと言にくい」という意見があった。

このような状況を踏まえて、要約筆記者の負担を減らし、聴覚障がい者に「即座に伝わる」、「複数の発言がわかる」、「聴覚障がい者が発言できる」ことを目指して開発したのが「LiveTalk」である。

3. LiveTalkの音声認識による文字化の特徴

LiveTalkとは、発話者の発言内容を音声認識エンジンで即座に文字化し、会議室内外の複数のパソコンに表示するソフトウェアである。LiveTalkの音声認識エンジンには、株式会社アドバンスト・メディア製のディープラーニングを使った音声認識エンジンを利用している（図1）。

なお、音声認識エンジンのコーパスとしては、日本語話し言葉コーパスを採用し、発話者が普段の発話で話しやすいスピードで1文を話しきるとすると認識率がよくなるため、発表者が事前にLiveTalkを使って練習することで、聴こえの問題がない者にとっても聞きやすいという副次的効果が得られ、聴覚障がい者への情報保証という以外にも情報を伝えるという本筋の品質向上も期待できるという特徴もある。

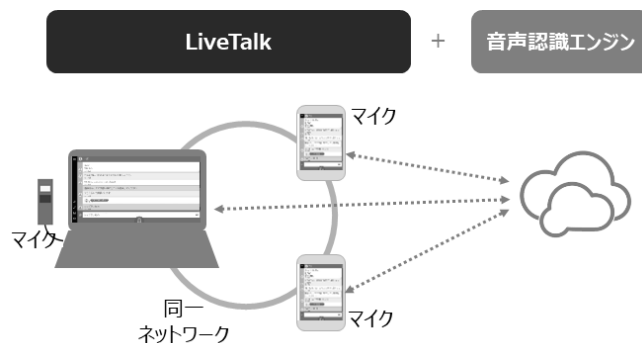


図1 LiveTalkの構成

3.1 即座に伝わる工夫

2013年当時の開発開始時点では、話し終わりまで話してから音声認識を行い、文字を表示する仕様が多かった。また、日本語音声認識でディープラーニングを用いた音声認識エンジンも少なく、現状のような認識

率ではなかったため、音声認識後に発話者が誤りを修正してから聴覚障がい者の前にあるノート PC に転送するという方式で PoC (Proof Of Concept : 概念実証) を実施した。しかし、聴覚障がい者からは「修正は後でもよく、発話が直ぐ表示されるリアルタイム性を重視する」というフィードバックが多く寄せられた。

そこで発話中でも音声認識エンジンから出力される認識結果をリアルタイムに表示する仕様へと変更を行って再度 PoC を実施した。その結果、リアルタイム音声認識版では次のようなメリット、デメリットが見えてきた。

表 1 リアルタイム音声認識での評価

	メリット	デメリット
リアルタイム性	どんどん表示されるので、リアルタイムに会話についていける	どんどん表示されるので、読むのが大変で会話についていけない
全発話表示	要約筆記と比較すると情報量が多く、いろいろ自分で考えることが出来る	要約筆記のように要約をしてもらった方が理解しやすい
誤認識	誤認識も確かにあるが、何を間違えたのか推測が付きやすいので問題ない	(聴覚障がい者当事者からの指摘なし)

特に、御認識に対する許容については想定をしていなかったし、質疑応答の時間に健常者から「誤認識が多くてまだまだ使い物にならない」とのご意見を頂いたが、聴覚障がい者の方から「誤認識があったとしても、文脈やそれまでの話から、何と誤認識したかが推測できるので、誤認識があるから導入しないという判断はしないでほしい」とのご意見をその場で頂き、情報が即座に伝わるということが聴覚障がい者の方にそれほど重要だということである。

リアルタイム性や全発話表示 (要約筆記との比較) でのデメリットについても、PoC を繰り返すことで、要約筆記に対する慣れからリアルタイム音声認識に対する慣れに移行する過渡期でのデメリットであり、要約表示方式を希望する声は上がらなく立ってきたため、製品版ではリアルタイム音声認識方式を選択した。

3.2 複数の発言の認識率を上げる工夫

聴覚障がい者が「会話にリアルタイムについていけない」要因の1つに複数人が同時に発話したり、発言中に誰かが割り込んだことで会話に変化したりしたときに気づけないということが挙げられる。

音声認識エンジンの特性としても、複数人が同時に

発話した場合、事前に発話分離などを実施して1人ずつ認識する必要がある。

そこで、発話者ごとにマイクを別にし、LiveTalk の画面上では、発話者ごとに表示枠を分け、複数の発話者が同時に発話したときには発話した順番に時系列に表示 (特許出願済) されるようにした (図 2)。

その結果、話の途中に割り込みがあったことがわかりやすくなり、聴覚障がい者へのヒアリングでも「複数の人が話していることが直ぐにわかる」と高評価であった。



3.3 聴覚障がい者が発言できる工夫

LiveTalk では、キーボード入力以外にもスタンプ入力や定型文入力など、聴覚障がい者が発言しやすい工夫を行っている。

当初は、キーボード入力のみを想定していたが、キーボード入力のスピードと他の参加者の発話スピードとでは差が大きく、参加者全員が聴覚障がい者のキーボード入力が終わるのを待たなくてはならなかった。そこで「困った」「わかった」「わからない」などを顔アイコンで示すスタンプや、前もって入力しておく定型文機能を設けた。これにより、例えば「おはよう」を前もって定型文として登録しておくことで、次回からワンタップで表示できるので迅速性が高まる。わからないことがあるときもスタンプですぐにヘルプを出せるようになり、コミュニケーションがスムーズになった。

4. 活用事例

LiveTalk の提供は当初ノートパソコンからスタートして、様々なプラットフォームに対応した。その中には、Microsoft HoloLens や EPSON MOVERIO などのウェアラブル機器も含まれている。以下に、特に反響が大きかったウェアラブル機器での活用事例を紹介する。

2017年9月、聴覚障がい者から「プラネタリウムの解説をリアルタイム字幕で同時に知りたい」というニーズにこたえ、実証実験を行った。通常、聴覚障がい者向けのプラネタリウム上映では字幕をコンテンツとして用意し、星空と同じ天球に字幕を表示するという方式が多く、健常者と別の日時に上映するというスタイルとなる。川崎市青少年科学館のプラネタリウム上映は、音声解説者がその場で解説するのが特徴であり、事前に表示する字幕を用意できないため、LiveTalk を使ってリアルタイム音声認識での文字による情報保証を実施した。

このとき、プラネタリウムの主役である星空をみることを阻害しないため、また、健常者と聴覚障がい者の親子が同じ時間に同じコンテンツを楽しめるようにする必要があった。そこで、星空に字幕を投影する方式ではなく、字幕が必要な来場者に EPSON MOVERIO (以下、MOVERIO) を貸出で、星空と一緒に字幕を見て頂くことにした。

また、プラネタリウムでの実証実験を開始する前に、黒壁を開発ルームに用意して MOVERIO に表示する文字の色などを検討した(表2)。その結果、背景が隠れないように透明背景にした状態では緑文字が好評であった。

プラネタリウム上映では、上記の検証結果に踏まえて文字色の明度を可能な限り落とした MOVERIO を貸出して、解説者の発話を音声認識した結果をリアルタイムに MOVERIO×10 台に配信し、星空の移り変わりや惑星間探査機の軌道などの解説を体感頂いた(図2)。



図2 プラネタリウムでの字幕表示

来場者からは、「星を見ながら解説も同時にみられたので理解が深まった」、「家族と一緒に見ることができた」といった好意的なコメントを得られた。特に多かったのは、字幕が不要な来場者と字幕が必要な来場者とが、一緒に一つのコンテンツを楽しめるメリットをあがる声であった。

5. 今後の展望

現在、音声認識エンジンはアドバンスト・メディア製の音声認識エンジン以外に Microsoft Azure Speech Services も選択可能であり、日本語以外の音声認識を使った文字による情報保証にも対応している。今後は、母国語の音声認識に強い地場の音声認識エンジンベンダーの製品も逐次採用し、日本だけではなく海外での聴覚障がい者への情報保証ツールとして活用できるように、富士通海外支社での取り扱いも開始した。

また、プラネタリウムの事例のように、健常者と聴覚障がい者が一緒にコンテンツを楽しめるような表示方法の提供も重要になってくるであろう。EPSON MOVERIO や Microsoft HoloLens などのウェアラブル端末への対応もその一環であり、今後も使いやすい新デバイスへの対応を積極的に実施していく。

コンピュータの入力方法は、紙テープ、マークカードやパンチカードから始まり、キーボード入力、マウス入力、フリック入力と人が自然な形で入力しやすく、入力速度も速い方式へと進んできている。すでに英語教材などでは存在しているが、日本語についても、スピーチ練習を音声認識で行って、苦手な発話をトレーニングするという学習方法も登場してくるだろう。

このように、今後は、音声認識エンジンの認識率のさらなる向上と共に、道具としての音声認識を自然に使いこなす利用者が増加し、誤認識も含めて、音声認識を使った文字によるリアルタイムな情報保証が一般化していくものだと考える。

6. まとめ

音声認識を使った文字による情報保証は、健常者が音声認識を使いこなす必要があるという構造ではあるが、一定レベル以上の音声認識率を実現することで、「使ってあげる」という他利的な目線から「使いたい」という自利的な自発利用での導入が増加する傾向にある。聴覚障がい者への情報保証の機会を増やすためには、健聴者にとって「リアルタイムに気持ちよく発話が文字化される」という価値観を追及していくことも重要な要素の1つである。

文 献

- [1] 富士通：聴覚障がい者参加型コミュニケーションツール「LiveTalk」
<http://www.fujitsu.com/jp/group/ssl/resources/news/press-releases/2015/0414.html>
- [2] 音声言語を文字言語化する「LiveTalk」×スマートグラス「MOVERIO」聴こえのバリアを乗り越える新たなプラネタリウム鑑賞の実証実験
<http://www.fujitsu.com/jp/group/ssl/resources/news/press-releases/2017/0822.html>