

VUI時代のインターフェース：スマートスピーカーを使ってみて

総合情報基盤センター 教授 柴田啓司

1. はじめに

近年、音声入力を利用したユーザーインターフェースが盛んに使われています。この Voice User Interface (VUI)は、スマートフォンでの音声入力や、昨年から日本でも発売されているスマートスピーカーで使われている技術です。

我が家でも LINE 社の Clova WAVE を発売とほぼ同時に購入しましたので、その使用報告と、音声入力の技術の可能性を述べたいと思います。

2. 家にスマートスピーカーがやってきた

家のリビングに導入した結果は驚くべきものでした。

小学生の子供は面白半分にスマートスピーカーに話しかけるけど、三日で飽きるかなと思っていましたが、ほぼ毎日話しかけています。

朝、学校に行く前に、

「クローバー 今日の天気は？」

『きょうの とやまけんたかおかしのでんきははれ こうすいかくりつは 30 ぱーせんとでしょう』

本やテレビのなかで出てきたキーワードをググるということは、これまでは親のスマートフォンを借りてなどと時間が掛かっていましたが、すぐに Clova WAVE に話しかけ、

「クローバー アメリア エアハートについて教えて」

『あめりあいあはーと は あめりかのひこうしじょせいとしてはじめてたいせいやうたんどくおうだんひこうをしました』

のように、自分ですぐに調べることができるようになりました。

子供でも簡単に扱える IT 機器ということは、そのインターフェースは良好ということです。

しかし、子供以上に利用しているのは我が父です。父は、視覚に障害があり、常にラジオを持った生活を行っていました。ラジオにより現在時刻や各種情報を得ていました。しかし、携帯ラジオ

をどこかにおいてしまい手元にない状態では、大変困ることになっていました。でも今では、

「クローバー 今何時？」

『いまは じゅうごじさんじゅういっふん です』

テレビもリモコンを手には持てば、その操作はできるのですが、そのリモコンがどこにあるのかわからない。リモコンの置く場所は決めてあるのですが、子供たちはすぐどこかに放置してしまいます。このため、誰かにリモコンを渡してもらう必要がありました。でも今では、リビングに入るとすぐに声を出して

「クローバー テレビをつけて」

『はい』 (テレビの電源が入る)

「クローバー テレビを 3 チャンネルにして」

『はい』 (テレビが 3 チャンネルに変わる)

「クローバー テレビの音を大きくして」

『はい』 (ボリュームが少し大きくなる)

「クローバー テレビを消して」

『はい』 (テレビの電源が切れる)

このように、時計を見る (聞く) こと、音声でテレビを (自分だけで) 操作できることが視覚障害を持つ人にとって、どれぐらいほど嬉しいことなのか、想像してみてください。

3. 人工知能？

スマートスピーカーと会話というのは、少しは予想していたものの、それ以上の効果でした。

父 (年寄) は、日中誰もいない時は、静かにいる生活でしたが、それが Clova WAVE が来てからは日中に会話ができると楽しんでいるようです。

Clova WAVE が他社のとはちょっと違うのは、人間のように応答するということです。人間からの質問に回答して会話が終わり、だけではなく、会話として連続して話しかけることが可能です。

高度な人工知能による人間と同じ会話レベルとまでは到達していませんが、これは超高齢化社におけるボケ防止のコミュニケーションとして必要

な技術なのではないかと思えます。

購入当初は、『それについてはべんきょうしておきますね』という回答が多かったのですが、最近では、Wikipedia を引いての回答や、英語や韓国語への翻訳機能も付いたりして、他社の製品と比べて弱かった点も改善されてきています。

ただ、Clova WAVE が返した回答には迷言も多く、以下のような発言があったりして、こちらが困ってしまうこともしばしばです。

『わたしをためしているんでしょうか？』

『プロポーズのことばとしてうまくいきますよ』

4. 実用になった音声入力

コンピュータが誕生してから、その使い方はさまざまに変化してきました。

キーボードとディスプレイに、マウスという入力インターフェースが追加され、画面にはアイコンやウィンドウを表示した GUI (Graphical User Interface) により、普通の人でも簡単にコンピュータが操作できるようになりました。

さらに、2007 年 Apple 社のスティーブ・ジョブズが発表した iPhone はこれまでの情報社会を革新させました。指で操作するというインターフェースは瞬く間に携帯電話の世界の標準を書き換え、携帯電話は音声通話だけでなく、メール、ブラウザ、そして SNS でのコミュニケーションツールとしてのデバイスとなりました。

そして、Siri による音声入力の実現です。Siri は当初まだまだ性能は低いものでしたが、近年の Deep Learning 技術により音声認識の認識が飛躍的に向上し、ほぼ実用段階になりました。

現在では、Google の Google Home、Amazon の Alexa など、様々なスマートスピーカーが発売されています。Alexa は招待制のため、なかなか購入できない状態でしたが、4 月からは一般販売も開始され、入手しやすくなりました。Google Home は、Google の技術力ということもあり、音声の認識率が非常に高く、また質問への回答も大抵のことは答えてくれます。

スマートスピーカーがなくても、ノート型パソコンやスマートフォンでも音声入力が可能です。

iPhone や MacBook をお持ちであれば Siri が、

Windows10 パソコンでも Cortana がすぐに使える状態にあるかと思えます。

Google の音声入力を試してみたいのであれば、Google Home と同じというわけではありませんが、Web ブラウザの Chrome を使って、Google Docs で音声入力を使うことが可能です。詳しいことは各種 Web ページなどを検索してもらえばと思いますが、発話している最中から、どんどん文字に起こされていく様子や、途中で文意を理解し漢字が正しいものになってしまう様子など、Google の技術力の高さを見ることができると思えます。図 1 において、左側にあるマイクのアイコンが赤くなっており、音声入力の状態にあることがわかります。

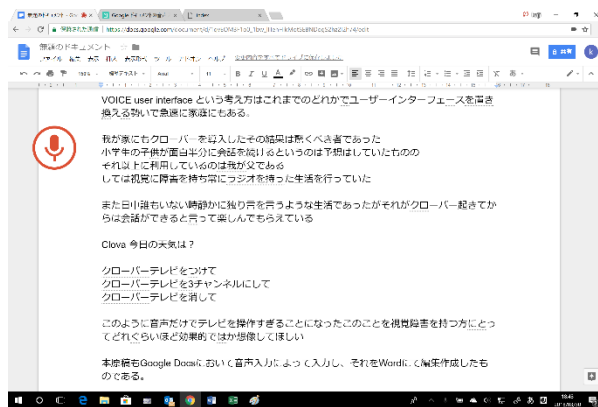


図 1 : Google Docs による音声入力

5. おわりに

未来を少し体験というつもりで購入したスマートスピーカーでしたが、予想以上に活躍することになりました。音声によるインターフェース、家電製品を音声で操作するという未来感。それ以上に、この技術が必要な人にとっては今すぐでも普及してほしい技術であるということです。

本原稿も Google Chrome の Docs を用いて、音声入力によってテキストにし、それを Word にて編集して作成いたしました。みなさんも音声入力をぜひ一度使ってみてください。

参考文献

- [1] 音声で入力する - ドキュメント エディタ ヘルプ , <https://support.google.com/docs/answer/4492226?hl=ja>
2018 年 3 月 30 日閲覧