

音声生成・音声認識 AI を用いた UI/UX デザイン教材の開発

辻 明典

徳島大学技術支援部

1. はじめに

AI 技術の急速な進展により、コンピュータへの音声やテキストによる入力、音声生成による出力の品質が格段に向上している。また、対話型 AI の高性能化により、コンピュータとのより自然な会話が実現され、コンピュータの存在を意識することなく所望の作業や動作を実現できる環境が整いつつある。本研究では、音声生成・音声認識 AI を実践する環境を構築し、AI 時代にふさわしい UI/UX デザインを考えるための教材開発を目的とする。人と地域共創センターにおいて開講中の公開講座「AI/IoT センサのしくみを知ろう」に本教材を導入したので報告する。

2. 音声生成・音声認識 AI の UI/UX デザイン

音声生成・音声認識 AI 技術の急速な発展を背景に、新たなコンピュータに求められる UI/UX デザインを検討するための教育教材を開発することを目的とする。近年、音声インターフェースが普及し、スマートフォンやスマートスピーカーに代表されるように、ユーザーは音声でデバイスを操作する機会が増加している。この流れの中で、UI/UX デザインは、従来の視覚的・触覚的なデザインに加え、聴覚的な要素を考慮した新たな設計が求められている。現存する UI/UX デザインには音声 AI に特化した内容は不足しており、音声入出力を主とした UI/UX デザインを学ぶための教材が少ないのが現状である。本研究では、音声生成・音声認識 AI の基礎から応用について、実践的に学ぶことができる教材の開発を行った。

3. 音声生成・音声認識 AI のシステム構成

音声生成・音声認識 AI を用いた様々なサービスが提供されている。しかし、その多くは有償サービスのため、演習や実習に用いるには、利用制

限やパソコンの性能不足、従量課金による支払いなどの課題がある。サービスの継続性、利便性を鑑みて、本研究では無償で利用できるオープンソースの音声生成・音声認識 AI をオンプレミスサーバ上に構築した。図 1 に本システムの概要を示す。ここで、音声認識 AI には Faster-Whisper、音声生成 AI には VALL-E-X を採用した。ユーザが自作のプログラムまたはコマンドから API 経由で音声認識・生成 AI にアクセスできるようフロントエンドの開発を行った。これにより、処理能力の低いパソコンやマイコンでも、音声生成・音声認識 AI を利用できる。

音声認識では、ユーザはマイクに話しかけて音声ファイル(WAV ファイル)を生成する。生成された音声ファイルは API 経由で音声認識 AI に送られ、結果としてテキストが得られる(文字起こし)。得られたテキストは、そのまま音声生成 AI に送り連携することもできる。音声生成では、テキストを API 経由で音声生成 AI に送り、結果として音声ファイル(MP3 ファイル)が得られる。また、同時にスピーカーで生成された音声ファイルを再生できる。音声生成 AI では、事前に固有の音声を用いて学習させて発話することもできる。ユーザインタフェースとして、従来のマイク、スピーカーが用いられるが、音声認識 AI、音声生成 AI との組み合わせによって、新たなサービスやユーザ体験を検討する開発環境を提供できる。

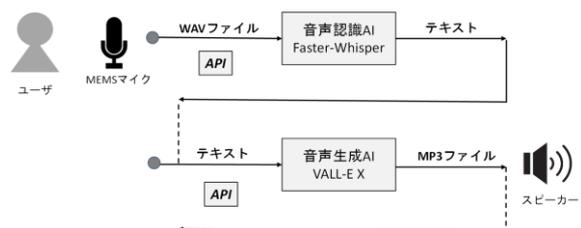


図1 音声認識・音声生成 AI を用いたシステム構成

4. UI/UX デザイン演習用教材

人と地域共創センターの公開講座「AI/IoT センサのしくみを知ろう」では、2023 年度より対話型 AI を用いたカリキュラムを導入して実施している^[1]。今回、音声生成・音声認識 AI を教材とするにあたり、表 1 に挙げる演習項目を追加した。

表 1 音声生成・音声認識 AI の演習内容

回数	演習内容
1	音声生成AI・音声認識AIの概要
2	音声認識AIと音声入力（マイク）
3	音声生成AIと音声出力（スピーカー）
4	APIを用いたコマンド・プログラムの実行
5	音声生成・音声認識AIの開発 1（I/O連携）
6	音声生成・音声認識AIの開発 2（センサ連携）
7	音声生成・音声認識AIの評価

演習内容において気を付けたのは、AI の利用がブラックボックスにならず、その原理やメカニズムを知りながら理解を深めるようにした点である。本講座では、徳島大学において開発した AI/IoT 向けの組み込みシステム教材を用いている^[2]。教材に実装された MEMS マイクを音声入力、スピーカーを音声出力として、音声生成・音声認識 AI にリクエストを送ることができる。また、教材にはスイッチや LED、ディスプレイなどの出力インタフェース、温度・湿度・気圧、空気質、人検知、GPS などのセンサ入力を備えている。音声入出力を主とした新しいコンピュータの UI/UX デザインを検討する上で、これら出力インタフェースやセンサ入力を「UI」の一つとして捉え、新たな視点での演習を行った。さらには、音声認識・音声生成 AI とこれら UI を連携させることで、現状の課題や将来的な実装、ユーザ体験「UX」の向上についての評価検討も行った。

5. 公開講座の実施と現状の課題

公開講座は春夏・秋冬の通年で実施され、春夏 10 回、秋冬 9 回(開講中)実施をした。受講者は、一般の方を対象とし、春夏・秋冬で 21 名の参加、40 代から 80 代の幅広い年齢層が受講した。受講者に、本教材についてディスカッションをする時間を設けてご意見・評価を頂いた。最も多かった

意見は、音声認識・音声生成の精度や正確さに関することであった。とくに、音声入力時、ノイズが多い環境や方言・イントネーションなどにより誤認識されることがある。また、音声認識 AI はサーバ処理の限界により、リアルタイムでの応答ができない課題があった。約 10 秒の音声認識に 6～7 秒の応答時間がかかるため、自然な対話を実現するに至らなかった。音声生成・音声認識 AI に限らず、一般的な生成 AI の特徴として、全く同じリクエスト(音声やテキスト)を入力しても実行結果がその都度変わってしまう。この点も、音声入出力の安定性に関わる部分だが継続の課題である。音声生成・音声認識 AI の API の提供により、気軽に AI に触れることができ良かったとの意見もあった。今後の音声認識 AI の精度の向上と合わせて音声入力の工夫が組み合わせると自然な対話に近づくことが期待できる。

6. まとめ

本稿では、音声生成・音声認識 AI を用いた UI/UX デザイン教材の開発について述べた。高度な音声生成・音声認識 AI の開発には多大なコストがかかるが、オープンソースを用いることで比較的 low コストで導入ができた。AI の開発競争は激しく、短期間の内に性能が向上すると考えられる。AI の応用は情報系に限らず、様々な分野での活用が期待されるため、最新の AI 技術を取り入れた教材開発は急務である。今後も、AI の仕組みの理解と応用展開、並びに UI/UX デザインの進展を見据えた取り組みを進めたいと考えている。

参考文献

- [1] 辻明典, “公開講座実習用教材への AI ペアプログラミングの導入”, 全学 FD 推進プログラム大学教育カンファレンス in 徳島, pp. 44-45, 2023.
- [2] 辻明典, 桑折範彦, 川上博, “AI/IoT オリジナル教材を用いた実験・演習プログラムの開発”, 全学 FD 推進プログラム大学教育カンファレンス in 徳島, pp. 46-47, 2022.