

研究報告

## 音声を用いたレトロフィットIoTの開発

本研究では情報ポートを持たない機器のIoTの手法として、音声認識と画像認識の2つの技術を融合したシステムを2例作成し、検討を進めました。その結果、両者技術がお互いの短所を補完することで片方の技術だけの場合よりも安定したシステムが構築できることが認められました。ここではその作成例の内、「薬品管理システム」の「音声入力」について紹介いたします。

### はじめに

ChatGPTの登場など、近年目覚ましく発展するAI技術ですが、そこではデータの存在が肝心であり、IoT技術に代表されるようなあらゆるデータを電子化し蓄積することがますます重要となっています。しかしながら、当センターの測定機器などにおいても十分な通信ポートを持たないものや、更には完全に機械的な装置で電子的な要素を持たない装置などがあります。このような機器のIoT化は「レトロフィットIoT」と言われ、重要な課題となっているところです。

一方で、これまで当センターでは製造現場における音声の活用、および画像認識の活用について検討を進めた結果、人の音声やジェスチャーによりある程度の情報入力が可能であることが分かりました。これら技術を駆使すれば、人が介在することにより、レトロフィットIoTを実現することが十分期待されると考えられます。ただ、音声認識においては外乱音に対する過認識が課題としてあった一方、画像認識については細かい判定を行う難しさから、入力できる情報数の制限が見込まれるところであり、それぞれに長所短所が認められます。

そこで本研究では人が介在したレトロフィットIoTの可能性を確認するべく、これら2つの技術を融合し、各々の短所を補完しあう形で、入出力を持たない機器に対しても活用可能なシステムの検討を進めました。

### システム構築について(音声認識と画像認識)

音声認識エンジンとしては、先行する研究と同様、無料利用可能な汎用大語彙連続音声認識エンジン「Julius」を使用しました。同エンジンは、①Linux / Windows / MacOSなど、対

応するプラットフォームの幅が広い。②エンジン起動後は音声でデータを入力するだけで解析結果が出力される。など、汎用性高く簡便に解析することが可能となっています。

一方、画像認識を行うモジュールについては、Alphabet社(Google)の「MediaPipe」を使用しました。MediaPipeにおいては物体の検出や追跡など、様々な機械学習ベースのソリューションが提供されており、特に人の特徴点(ランドマーク)を取得するソリューションが豊富です。本研究では、「顔のランドマーク取得(FaceMesh)」、「手のランドマーク取得(Hands)」、そして「体のランドマーク取得(Pose)」の3つを適宜用いることとしました。

### 試作システムについて

この度試作したシステムの1つが、当センターで運用している「薬品管理システム」(図1左図)の音声認識システムです。「薬品管理システム」はQRコードを用いて、使用者や薬品瓶種などを端末に読み込ませた後、天秤で計量し、そのデータを通信線で端末に転送。それをデータベースとして登録していくシステムとなっています。ただ先述のとおり、当センターの天秤はアナログ的なものもまだあり、十分な通信ポートを持たないものも多くあります。そこでこの度、このアナログな天秤でもデータベースへの登録を可能にするため、画像認識の補助を用いつつ音声での

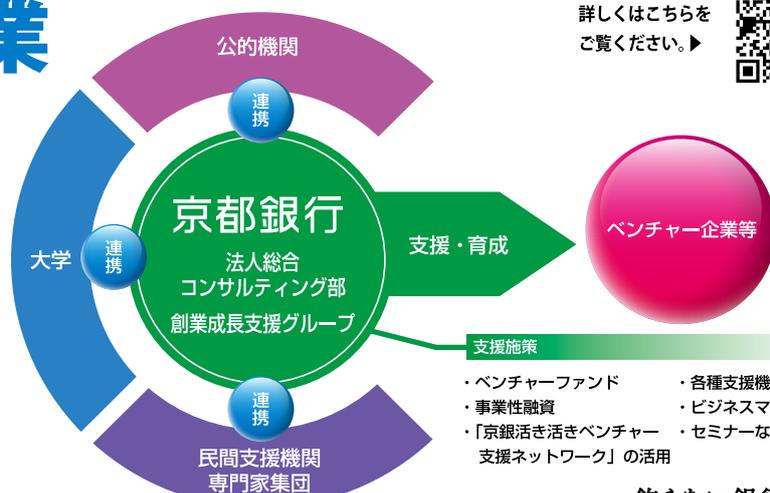


図1 薬品管理システム

# ベンチャー企業 支援業務の ご案内

#### 業務内容

- ベンチャーファンドによる株式投資やご融資を通じて、事業資金のサポートを行います。
- 資金面の支援だけでなくとどまらず、公的機関・専門機関・大学等のネットワークである「京銀活き活きベンチャー支援ネットワーク」等を通じ、経営相談をはじめベンチャー企業のあらゆるニーズにお応えします。



詳しくはこちらを  
ご覧ください。▶



飾らない銀行  
**京都銀行**

入力ができるようなシステムを試作しました。

なお、もう一つは試験装置の遠隔操作にかかるシステムですが、その仔細は当センター技報No.51をご覧くださいいただけます。

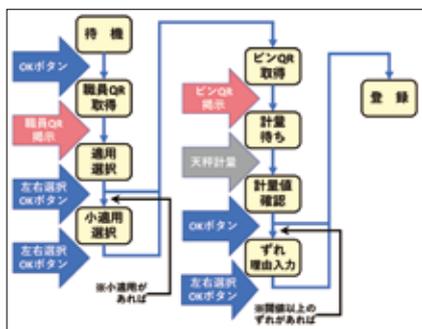


図2 薬品管理システムのフロー図(例)

## 薬品管理システム

図2は、薬品管理システムの作業フローチャートです。このように適宜、システムのボタンを押し項目を選択し、QRコードを読み込ませて各種情報を入力する流れとなっています。そして重量については天秤の計量ボタンを押すことで、端末へ転送・取得し、最後にデータベースへ登録するという流れになります。この度は ①計量結果を数字で読み上げることで端末に入力すること ②音声認識やデータ登録についてはジェスチャーで認識させることを検討し、さらに ③QRコードを使わない手法として、各種選択項目についてもジェスチャーと音声での対応を検討しました。

## 試作結果

図3は、この度作成した音声認識対応薬品管理システムユーザーインターフェース画面です。入力項目は、薬品の場所、天秤種、利用者、どの目的での重量計測か(所作)、種別(瓶)、重量としています。これらの項目は、親指と人差し指を伸ばしたジェスチャー(図4(a), (b))により昇順・降順で繰ることが可能です。また音声認識の際は、親指と小指を伸ばしたジェスチャー(図4(c))をトリガとして認識を開始します。この2つのジェスチャーで項目を選択し、データを読み上げ入力するということを実現しました。

またデータベース登録という不可逆的な操作に関しては、2つの手を使うジェスチャーで実装しており(図4(d))、この状態で5秒間、認識させ続けることとしています。この度、数字の入力に関しては数字を1つずつ読み上げる実装方法を用いました。時々間違えた認識をするものの、改めて正確に読み上げなおすと、しっかりと認識されました。なお入力項目をジェスチャーで繰ることを採用しましたが、この作業は煩雑な面もあったため音声で認識項目を指定する方が望ましいと思われる、どのような入力が音声であるべきか、画像であるべきかについては十分に検討することが重要と考えられます。



図3 ユーザーインターフェース画面



図4 音声認識のジェスチャー画面

## まとめ

本研究では通信ポートを持たない機器のIoT化というレトロフィットIoT背景から、音声と画像を用いた入出力方法を検討し、ここではその成果の1つである薬品管理システムの音声認識システムを紹介しました。本研究の結果から、音声と画像の2つを組み合わせることでより堅実で利便性の高いシステムが組められると思われます。ただ、音声と画像のどちらにどのような役割を持たせるかはよく検討する必要があり、使用環境の外乱等々も検討しながらケースバーケースでの対応が必要とも思われます。

本研究のみならずこのようなIoTの取り組みにご興味がある方は、お気軽にご相談頂ければ幸いです。

●お問い合わせ先/ 京都府中小企業技術センター 応用技術課 電気通信係 TEL:075-315-8634 E-mail:ouyou@kptc.jp

— 想いをむすび、地域をゆたかに —

# 京都信用金庫

京信 ソーシャル・グッド預金

Kyoshin Social Good Deposit

“世の中を少しでも良くしたい”という預金者の想いを6つのテーマに乗せて企業に託し、今も未来も安心して過ごせる地域をともに創るための預金です。

ソーシャル企業認証制度 S認証

ESG経営や社会課題の解決を目指す企業の評価・認証を行い、**企業活動の社会的インパクトをみえる化する**制度です。社会課題に取り組む地域企業の成長を支えます。

京都信用金庫は、地域社会におけるソーシャルマインドの醸成及び持続可能な地域社会の実現を目指します。