

# 端末型音声認識を用いた現場作業支援システムの実用性検討

渡辺奈夕子<sup>†1</sup> 藤村浩司<sup>†1</sup>

**概要:** 保守点検などの作業現場ではクラウド型音声認識や対話サービス等が使われ始めている。しかし、実際の現場ではネットワーク環境が貧弱な可能性があり、また作業スピードを重視しシステムの応答が速いことが求められるため、クラウド型のサービスや対話サービスの利用が難しい場合がある。また、対話サービスに沿って返答を入力する形式の場合、作業者が好きな順番・タイミングで入力を行うことができない。そこで、スマートフォンなどの携帯端末で動作する音声認識を用いて点検結果や作業結果の入力を行う、現場作業支援システムを提案する。提案システムでは作業現場ですぐに運用可能な支援システムを目指しており、作業者が自分の慣れた順番やタイミングで発話、入力を行うことができる。本発表では、提案システムの紹介と、実用性の検証を行った結果について報告する。

**キーワード:** 音声認識, 保守点検, 作業支援

## A Study of Practicality for Field Operation Support System using Speech Recognition on Handheld Devices.

NAYUKO WATANABE<sup>†1</sup> HIROSHI FUJIMURA<sup>†1</sup>

**Abstract:** Cloud speech recognition services and speech dialogue services are used for many systems supporting various maintenance works. However, cloud services are sometimes inappropriate because wireless network is often unavailable on actual work area, and the system must quickly respond to operators. Furthermore, simple dialogue services force the workers to do their tasks in predetermined order. We propose a field operation support system using speech recognition running on handheld devices. Our system enables workers to enter various data to the system by speech in their usual order. In this paper, we show our system in detail and experimental results to verify practicality of our system.

**Keywords:** Speech Recognition, Maintenance Work, Field Operation Support

### 1. はじめに

クラウド型音声認識や AI を用いた対話サービスが一般的になり、広くユーザに使われるようになってきている。また、社会インフラなどの保守点検の作業現場でも、音声認識や対話サービスを用いて作業支援が使われるようになってきている。例として、スマートフォンやタブレットなどの携帯端末から作業内容・点検内容を合成音声で指示し、それに対する作業者の作業結果・点検結果に関する発話内容を音声認識することで、作業支援を行うと共に作業結果のデータ化を行うサービスが多々ある[1][2]。作業エビデンスを残す、作業を見える化し業務効率の改善を行う等の目的のため、このような音声認識等のサービスを使った作業自体のデジタル化は重要である。

しかし実際の作業現場では、ネットワーク環境が貧弱でクラウド型のサービスが適さない場合がある。例えば客先の地下室での点検作業などでは、ネットワーク環境を保証することはできない。さらに保守点検は出先での作業が連続することがあり、省電力であることが重要となる。そのため、ネットワーク越しの処理や高負荷な処理はできるだけ少なくしたい。

また一般的なユーザの用途と違い作業の速度を重要視するため、一つ発話をする度にシステムからの応答や指示

を待つのは現実的ではない。特に習熟した作業者では、システムの応答や指示よりも早く作業を進めることができ、システムの反応を待つことは逆に作業者の負担になりかねない。また習熟した作業者には慣れた作業手順があり、システムの指示順通りに作業をさせることも負担になる可能性がある。

そこで本稿では、携帯端末単体で動作する軽量なトリガ検出機能とコマンド型の音声認識を用いた現場作業支援システムを提案する。提案するシステムではシステム側からの音声での指示は行わず、作業者が自分のペースで作業結果や点検結果の複数の項目に対して音声入力を行うことができる。また本稿では提案システムの実用性について、評価実験を通して検証した結果を報告する。

### 2. 関連研究

音声認識や音声合成を使って現場の作業支援を行うサービスは多くある。項目に対する値の入力に音声入力を使う例[1][2][3]では、合成音声で入力する項目が指定された後、作業者が該当する項目値を発話すると、発話内容が音声認識されて項目値がデータ上に入力される。もしくは作業内容が合成音声で指示され、それに対して作業者が実際に作業を行い、結果を音声で入力する、という流れになっている。この方法では、システムが入力対象項目を絞って

いるため、その後に発話する項目値の音声認識の際に項目に適した音声認識辞書を用いることができる。そのため、認識項目を絞らない場合に対して、認識精度が良くなるという利点がある。例えば「型番」などの項目では英数字の羅列を認識する必要があるが、「電圧」であれば数値の認識をすれば良いし、点検結果の良否の入力だけの場合は「OK」「NG」など、限られた単語のみを認識すれば良い。一方、システムからの指示に対して作業や入力を行う必要があるため、作業者が自分の慣れた順番や速度で作業をすることは難しい。

連続音声認識を用いて作業者の自由発話を認識し、その中から複数の項目に入力すべき内容を抽出する例[4]もある。この場合は前述したようにシステムの指示を待つ必要は無いが、どの項目の認識にも同じ音声認識辞書を用いることとなる。特に作業現場では型番など言語的に意味が無い英数字の羅列を認識しなければならないことが多く、一般的な連続音声認識では認識精度が落ちてしまうという問題がある。

### 3. トリガ検出とコマンド型音声認識を用いた作業支援システム

本システムは関連研究と同じく、音声を用いて点検結果や作業結果などの複数の項目の入力を行うことを目的としている。前述した通り、音声での作業結果の入力は入力対象の項目を指定した上でそれに対する値を発話するものが多く、作業者は自分の慣れた順番、タイミングで作業することができない。対して本システムでは作業者が項目名も含めて発話することで、システムの指示を待つことなく、作業者の慣れた順番で作業を行うことができる。

#### 3.1 対象とする作業例

現場の作業には様々な種類がある。例えば機械や施設の保守点検作業であれば、点検を行いながら点検項目に点検結果を入力していく。点検結果が異常値であれば作業者に知らせると共に、結果を記録していくことで結果報告書の出力や作業効率の改善につなげることができる。また他にも、定められた作業が正しく行われたかどうかをチェックする作業や、部品などが正しく規格通りの性能を示すか否かを確認する作業もある。

本稿では複数の項目を入力する例として、複数の品物をグループ毎に仕分ける作業が正しくデータ通りに行われたか否かを確認する、検査作業を用いて説明する。従来この検査作業は、2名の作業者のうち一方が現物を見て項目値を読み上げ、もう一方がデータを出力した紙を見ながら読み上げを聞いて、対象の項目を探し出し、読み上げがデータと合致するかどうかを照合し、紙の上で照合結果を書き込んでいた(図1)。

これに対し提案システムでは、検査作業の“紙のデータを見ながら読み上げを聞いて正しいかどうかを照合する”

部分をシステムが担う(図2)。作業者はヘッドセットを装着し、現物を見ながら項目名と項目値(「型番 RS8Q」など)を発話する。システムは読み上げ内容を音声認識し、読み上げられた項目(型番)の正常値のデータ「RS8Q」と認識結果を照合し、合否のフィードバックを音と画面表示で行うと共に、それを記録する。

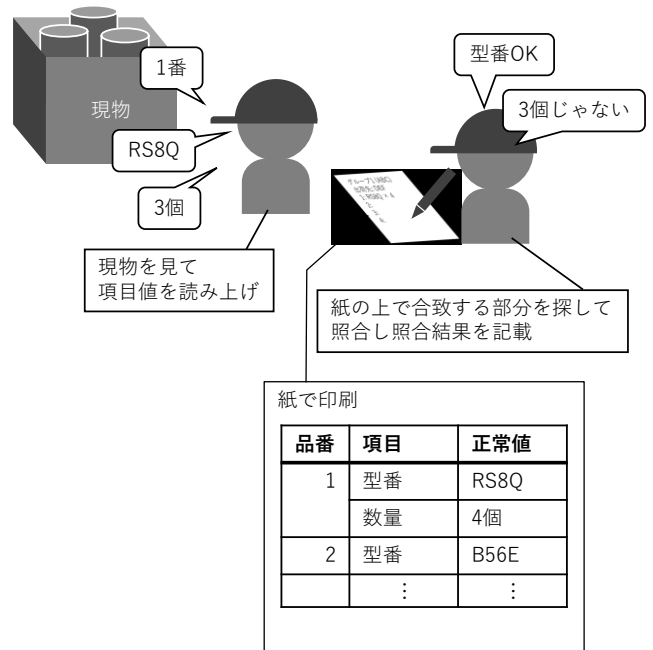


図1 従来の検査作業

Figure 1 Operation by two workers with printed data.

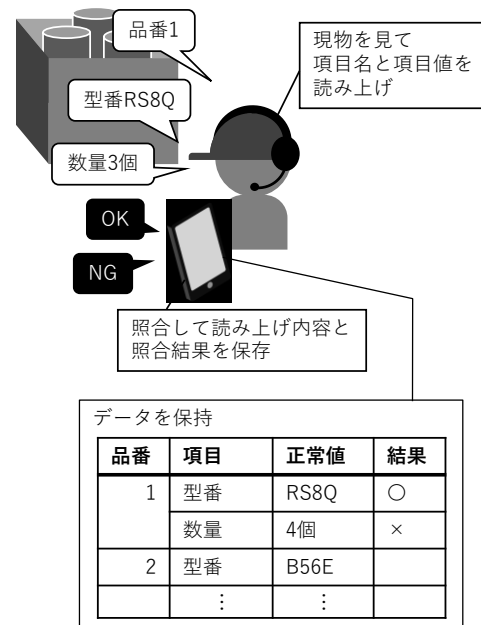


図2 提案システムを使った検査作業

Figure 2 Operation using our system.

表1にこの作業設定の上で入力する内容の例を示す。こ

の作業設定では、入力内容は複数のグループに分かれており、グループには“グループ ID”と“出荷先”という項目がある。また各グループに複数の“品番”があり、品番ごとに“型番”、“品名”、“数量”、“色ラベル”の項目がある。

表 1 入力する内容の例。灰色の部分に項目名を表す。

Table 1 Example of input data. Item names are represented as hatched.

グループ ID		TT74K			
出荷先		青森			
品番	型番	品名	数量	色ラベル	
1	RS8Q	CD ドライブ	1	黄	
2	GXO	水冷パーツ	1	黄	
5	2M2	SSD	7	灰	
グループ ID		NHX8			
出荷先		福島			
品番	型番	品名	数量	色ラベル	
2	CNB	DVD ドライブ	8	黒	
3	IRIC	サウンドカード	4	赤	
⋮	⋮	⋮	⋮	⋮	

### 3.2 ユーザインタフェース

図 3 に、提案システムの画面例を示す。グループ番号を指定すると (図 3(a)), 指定したグループが選択された状態で一覧画面が表示される (図 3(b)の明るい部分)。この時点で音声認識が開始され、指定したグループの情報を入力可能になる。作業者は検査対象の現物を見て各情報を読み上げる。



図 3 提案システムの画面

Figure 3 User interface of our proposed system.

例えば項目名“グループ ID”の値“TT74K”を入力する場合、「グループ ID」と発話すると、システムはグループ ID の入力モードに入ったことを「ピコン」音と画面上でフ

ィードバックする (図 4(a)). 続いて作業者が「TT74K」と発話すると、システムはその認識結果を内部に保持しているデータと照合し、合致すれば「ピンポン」音を再生すると共に、正しい結果であったことを画面上で示す (図 4(b)). 合致しない場合は「ブブー」音を再生すると共に誤った結果であったことを画面上で示す (図 4(c)). グループ ID の入力モードに入ったことを確認してからその後の値を発話していると時間がかかってしまうため、入力モードに入ったフィードバックを待たずに値を発話することができるようにしている。つまり「グループ ID (ピコン音待ち) TT74K」と発話しても、「グループ IDTT74K」と発話しても良い。



図 4 “グループ ID”入力時の画面

Figure 4 Interface & feedback for input “グループ ID”.

“品番”の各項目を入力するには、「品番 1」のように発話し、1 番の品番の入力モードに移行する (図 5(a)). その後は「型番 RS8Q」など、部品の各項目についてグループ ID 等と同様に入力を行う (図 5(b), 図 5(c)). 「品番 2」のように別の品番を発話すると、その品番の入力モードに移行する。また、品番の入力モード中に「グループ ID」や「出荷先」と発話すると、即座に品番の入力モードを抜けて指定した項目の入力モードに移り、項目値の発話が可能となる。

項目値の発話が誤認識された場合は、再度項目名から発話することで入力し直すことができる。または、各項目をタップするとその項目の音声認識結果と入力ダイアログが表示され、その上で手入力にて修正することができる (図 6).

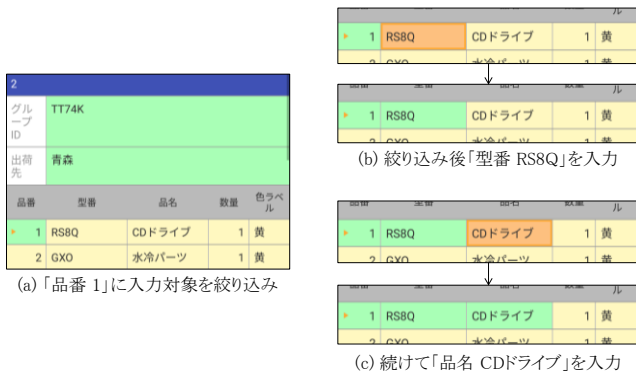


図 5 品番の各項目入力時の画面

Figure 5 Interface & feedback for input items of '品番'.

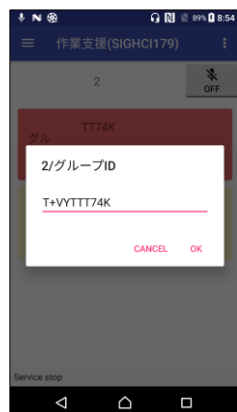


図 6 手入力での修正画面

Figure 6 User interface to correct a recognition result.

## 4. 実装

本章では提案システムの実装について述べる。

提案システムは、Android™ OS が動作するスマートフォン上のアプリケーションとして実装した。

### 4.1 トリガ検出とコマンド型音声認識

Google Home[5] や Amazon Echo[6] などの AI スピーカーなどで使用されているトリガ検出機能（「OK Google」や「Alexa」などのトリガワード、ウェイクアップワードを検出する機能）は、スマートフォン等の携帯端末でも低コストで利用可能である。トリガ検出は予め定められた少数のトリガワードが音声に含まれているか否かを検出する技術である。低消費電力のトリガ検出を常に動作させておき、トリガワードが検出された場合のみ音声認識を起動することで、不必要な認識結果を得ることが無く、必要なときのみ認識することができる。

一般的な発話文を認識する連続音声認識を携帯端末で動作させることは処理速度等の問題があり難しいが、制限された文法のみを認識できるコマンド型音声認識は、携帯端末で動作させることが可能である。また、型番のような英数字の羅列を、一般的な文章を認識するための連続音声

認識で認識するのは非常に難しい。

提案システムでは、作業中常に音声認識を動作させると負荷が高く、また不要な語を認識してしまうため、一般的な AI スピーカーと同様トリガ検出を用いて、コマンド型音声認識をする範囲を限定する。しかし作業中、入力の度に特別なトリガワードを発話させることはスムーズな作業を妨げてしまう。そこで、「グループ ID TT74K」のような読み上げの、項目名部分である「グループ ID」の部分トリガワードとして用いる。項目名がトリガワードとして検出されたら、その後の発話について検出された項目に応じた音声認識辞書を用いたコマンド型音声認識でその後の「TT74K」部分を認識する。

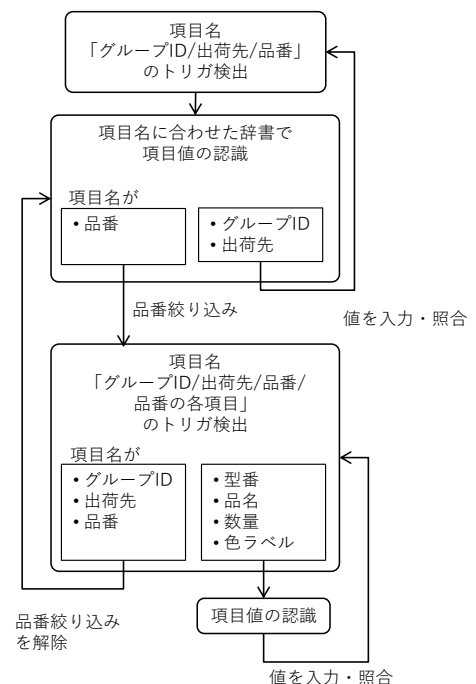


図 7 提案システムの処理フロー

Figure 7 Process flow of the proposed system.

実際の音声認識制御のフローを図 7 に示す。最初は「グループ ID」「出荷先」「品番」のトリガワードをトリガ検出で待ち受ける。検出された項目名に合わせた辞書を用いて、その後に発話される項目値の内容をコマンド型音声認識で認識する。検出された項目名が「グループ ID」・「出荷先」の場合は、音声認識結果を DB 上に入力し、データとの照合を行う。項目名が「品番」の場合は品番を絞り込んだモードに切り替える。品番を絞り込んだ状態では、「グループ ID」「出荷先」「品番」と、品番の項目「型番」「品名」「数量」「色ラベル」のトリガワードをトリガ検出で待ち受ける。「グループ ID」・「出荷先」・「品番」が検出されたら、絞り込みを解除したうえで項目値の認識を行う。それ以外の項目名が検出されたら、検出された項目名に合わせた辞書を用いて項目値の認識を行い、認識結果の DB への入力、デ

ータとの照合を行う。

#### 4.2 音声認識結果とデータの照合

コマンド型音声認識結果は、その信頼度の順に複数の認識結果候補で得られる。認識する対象が少ない（OK と NG のみなど）場合には認識結果候補は 1 つで良いが、英数字の羅列など認識精度が低下しやすい内容を認識する場合には、複数の認識結果候補の中からデータと一番近いものを選択して照合を行う。

数字の“9”とアルファベットの“Q”のように読みでは区別できない文字や、数字の“0”とアルファベットの“O”のように現物を見た際に区別がつかず更に双方とも「マル」のように読む可能性のある文字などは、曖昧性を許容してどちらが認識されてもデータと合致したとみなす。また、記号の読み方は人によって違う上に、記号を読まない場合も考慮し、記号の有無がデータの違いを表す場合（例えば品番に“1 (+)”と“1 (-)”の両方がある場合など）を除き、記号の有無についても無視して照合を行う。

### 5. 実用性の検証

提案システムの実用性を検証するため、3.1 で述べた模擬的な検査作業を行う評価実験を実施した。実験では画面サイズ 6 インチで Android 7.0 のスマートフォンを用いた。

#### 5.1 被験者

被験者は音声認識に携わる研究者 9 名（うち女性 2 名、男性 7 名）で、全員が日本語話者である。

#### 5.2 タスク

被験者は PC 上で表示される表 1 のような入力する内容の表（現物を表す）を見て、各項目の入力を行う。また途中でスマートフォン端末の画面を見ても良い。実際の作業では表が誤っている場合がありそれをチェックするための作業であるが、本実験では与えられた表は正しいもの（端末内に保持するデータと合致したもの）を用いる。このため被験者が正しく発話を行い、トリガ検出の結果と認識結果が正しい場合はデータと合致するはずである。被験者は照合結果が合致するまで入力を行う。項目名の検出がされなかった場合、項目の値が誤認識されてデータと合致しなかった場合、誤って発話してしまった場合は、もう一度項目名と項目値を入力し直すこととした。

5 個の品番があるグループを、1 グループ分入力することを 1 タスクとし、1 人あたり練習用 1 タスク、本番 5 タスク入力してもらった。評価に用いた 1 人あたりの入力項目は、5 タスク×(グループ ID+出荷先+5×品番項目 5 個)の 135 項目となる。実用できるかどうかを調べるため、全入力にかかった時間や、照合結果が不整合の場合の不整合の内容等のログを取った。また全入力が終了した後、使い勝手等についての意見を聞いた。

また上記タスクとは別に、提案システムを使用せずに 2 人作業で行った場合にかかる時間との比較をするために、

2 人のうち一方が評価実験と同様に 5 グループ分の読み上げを行い、もう一方が紙に印刷したデータ上にチェックをつける、という作業を 1 組行なった。

#### 5.3 タスクで入力する内容

入力項目ごとに受け付ける項目値の内容を表 2 に示す。例えば“グループ ID”であれば英数字や記号が羅列する文字列を入力できる。“色ラベル”は項目値の内容が色のリストなので、予め定められた色の集合の中から 1 つを入力する。“品番”は入力された項目値を照合するのではなく、注目している（その後発話する項目が属している）品番を絞り込むために使用する。

入力項目は全項目が表 2 のルールに沿ってランダム生成したものをを用いた。

表 2 項目値の規則

Table 2 Rules of item values.

項目名	項目値の内容
グループ ID	英数字+記号 (.-()/+@) 5 文字~10 文字
出荷先	出荷先のリスト (48 個)
品番	数字 (1~9)
型番	英数字+記号 (.-()) 3 文字~5 文字
品名	品名のリスト (18 個)
数量	数字 (1~9)
色ラベル	色のリスト (9 個)

### 5.4 結果と考察

#### 5.4.1 作業にかかった時間

全タスク (5 グループ分の入力) の作業にかかった時間は、被験者平均で 715.4 秒であった。これに対し、2 人で行った場合の作業にかかった時間は 307 秒であった。図 8 および表 3 に、入力項目ごとの、1 項目あたりの入力にかかった時間の平均を示す。

表 3 1 項目の入力にかかった時間の平均 (秒)

Table 3 Average time in seconds of each item.

入力項目	提案手法	2 人作業
グループ ID	31.51	6.82
出荷先	4.98	2.24
品番	3.87	1.97
型番	5.65	2.98
品名	4.38	1.93
数量	3.68	1.56
色ラベル	3.75	2.06
平均	5.30	2.28

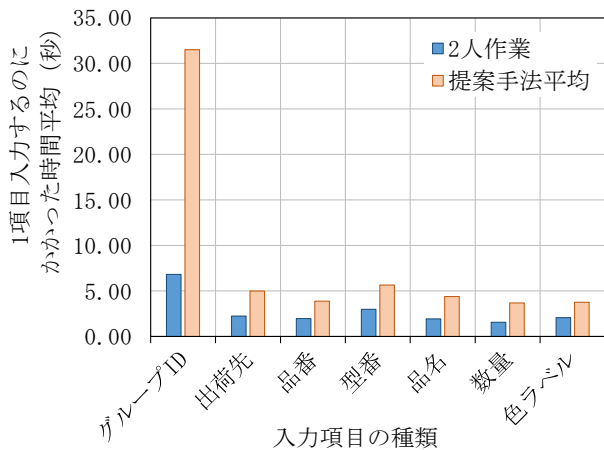


図 8 1 項目の入力にかかった時間の平均 (秒)  
Figure 8 Average time of each item (sec).

提案システムを使った場合、2 人作業にかかった時間の 2 倍以上の時間がかかっているため、工数の削減にはなっていない。入力項目の種類別に見ると、特にグループ ID の入力に時間がかかっていることがわかる。他の項目では 2 人作業に対して 1.8~2.4 倍の時間で入力ができているのに対し、グループ ID では 4.6 倍の時間がかかっている。グループ ID 以外の項目では工数こそ削減できないものの代替可能な範囲であり、デジタル化をするという観点で言えば実用可能であると思われる。

グループ ID の入力に時間がかかってしまった要因は、グループ ID において入力の失敗が多かったこと、つまりトリガ検出におけるトリガワードの未検出（発話したのに検出されない）や誤検出（発話したワードと違うものが検出される）、コマンド型音声認識での誤認識が多かったことである。

表 4 1 項目の入力が成功するまでに失敗した入力回数

Table 4 Times of input errors.

入力項目	1 項目あたりの失敗回数
グループ ID	1.18
出荷先	0.09
品番	0.05
型番	0.26
品名	0.07
数量	0.11
色ラベル	0.06
平均	0.15

表 4 に、1 項目の入力が成功するまでに行った失敗入力の回数を示す。グループ ID と型番については英数字の羅列の入力であり、音声認識を行うのが非常に難しい。その

ためグループ ID と型番の失敗回数は他の項目よりも多い。グループ ID の失敗回数は 1 を超えており、これはグループ ID を入力する際に 1 回は入力に失敗してしまう、ということを表す。

### 5.4.2 作業の習熟による変化

タスク順に、1 グループ分を入力するためにかかった時間の被験者平均を図 9 に示す。また、2 人作業における 1 グループあたりの入力時間も併せて示す。

タスクが進むにつれ、入力にかかる時間が 2 人作業のおよそ 2 倍程度になることがわかる。グループ ID の時間が長かったのが段々短くなり、入力項目ごとの比率が 2 人作業に近づく。つまり被験者はグループが進むにつれて提案システムでの音声入力に習熟していったと考えられる。

英数字の羅列の入力は慣れないと正しく音声認識されないことが多い。今回の被験者は全員が音声認識の研究開発に携わっているため英数字の羅列の習熟が早かった。一般の作業者ではもう少し習熟まで時間がかかると思われるが、習熟することでグループ ID のような長い英数字の羅列入力でも実用可能となりえる。

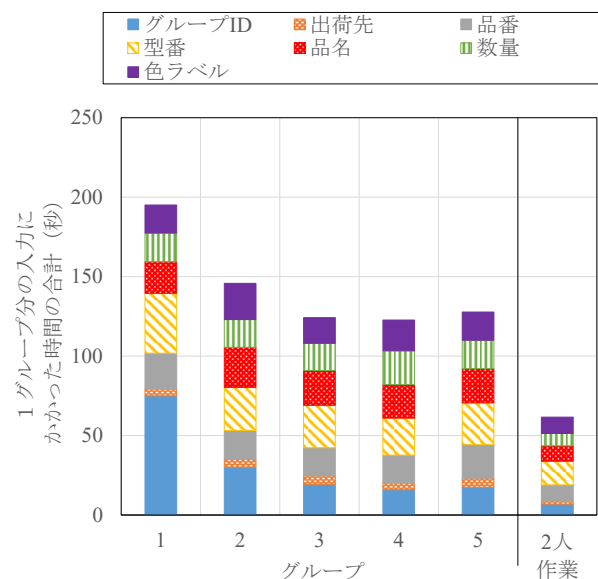


図 9 1 グループ入力するためにかかった時間の被験者平均 (秒)

Figure 9 Average times of trials.

### 5.4.3 得られた意見

提案システムを使った入力について、被験者に使い勝手等について意見を貰った。

#### (1) フィードバックについて

照合結果が合致しなかった場合には音のフィードバックとして「ブー」音を再生していたが、このフィードバックだけではそもそも音声認識が間違っていたのか、読み上げたデータ（現物）が間違っていたのかの区別がつかない。また誤認識の場合、どこが間違っていたのかが作業者



にはわからず、何度発話しても同じように誤認識してしまう、という意見があった。

項目名と項目値の間を空けずに発話しても良く、その場合でも項目名が検出された際に音のフィードバックがある。すると発話している間に項目名が検出された音が鳴ることになり、煩わしいという意見があった。逆にフィードバックが無い場合には、認識に時間がかかっているのかと思い（実際には項目名が検出されていない）、いつまでも待つてしまうとの意見もあった。

## (2) 入力について

グループ ID のように長い英数字の羅列は一気に読むのが難しく、また、どこで切って発話して良いのかわからないとの意見があった。一般的な型番では“-（ハイフン）”などの記号が挟まれていることも多いため、その場合には大きな問題にならない。そうでない場合には発話を始めてから悩むと言いつつ間違えてしまうので、データの種類などに応じて切る場所を予め決めておくなど、運用でカバーする必要がある。

今回英数字の読み方については特に指定をしていないが、コマンド型音声認識の辞書で対応していない読み方をしてしまうと認識することができない。これについて、どの程度の読み方のバリエーションで発話して良いのかわからないという意見があった。例えば“0”は作業現場では「マル」と読むことも多い。また、色を英語で読んで良いのかと実験中に質問されることもあった。また英語で読むつもりは無くても、英数字を呼んでいる間に数字を英語で読んでしまうことや、英字の羅列が英単語になっている場合にそれを英語読みしてしまうことがあった。これらについては音声認識辞書の読みのバリエーションを多くしておくことで解決するが、認識対象が多くなる程精度が悪くなるため、システムの運用時に読み方についてはある程度制限しなければならない。ただし制限すると作業者の作業の自由度を狭めることになるため、トレードオフである。

項目名と項目値の間を空けずに発話することについては、いちいち待たなくて良いので楽という意見があったと同時に、間をどの程度空けて良いのかわからない、という意見もあった。また、照合時のフィードバック音を待たなければその次の項目の入力はできないが、そこも間を空けずに複数項目を一気に入力できると嬉しい、との意見があった。

## 5.5 ユーザインタフェースの改良案

以上を踏まえ、今後実用性を高めるために必要となるユーザインタフェースの改良について考察する。

今回の作業では正解データがわかっているため、それと合致しない場合に何かしらのフィードバックを返せば良いが、実際の作業では、誤認識が起きたのかそれとも実際に読み上げた内容（つまり現物）が誤っているのかの区別がつくようにしなければならない。誤認識なのか現物の間違い

いなのかを区別するには、認識結果と正解データの類似度や不整合の内容、認識結果に付随する信頼度の情報、入力回数（同じ項目を何度も入力し直しているかどうか）などを用いることができる。例えば不整合の箇所が誤認識しやすい間違いの場合は誤認識の可能性が高い。その場合は認識結果と正解データの違う個所を強調して画面上で表示することで、どこが誤認識されたかを作業者が把握しやすくなる。例えば英文字「H」は「A1」などと誤認識しやすい。正解が「30M2HKD」で認識結果が「30M2A1KD」になってしまった場合は、画面上でのように表示するとともに、音声合成で H が A1 になっている旨をフィードバックすればよい。また正解データと認識結果が近い場合はその箇所だけ手入力や誤りやすい内容（この例の場合は H）の選択をできるようにすることで、手入力の手間も削減できる。誤認識をしやすい箇所（文字）は人によって違うため、音声認識の個人適応と共に、認識結果のフィードバック時、手入力画面表示時の個人適応も有効であると考えられる。

2	
グループ ID	入: 30M2A1KD 正: 30M2HKD
出荷先	青森

図 10 認識結果と正解データの不整合の表示

Figure 10 User interface that shows audio recognition result and expected value.

また正解データが無い作業の場合（例えば電圧などのメーターの値を入力するなど）は、そもそも誤認識されたことに気付く方法がない。この場合は、認識結果を音声合成で読み上げてフィードバックする方法がある。この場合でも認識結果の信頼度は利用でき、信頼度が低い場合は作業者に聞き直して再度入力を促しても良いと考えられる。ただし音声合成でのフィードバックが多くなると作業のテンポが悪くなるという問題が残るため、何度も間違えた場合のみ認識結果を読み上げる等の工夫は必要である。

これらの方法をとっても、グループ ID のような長い英数字の羅列は入力が難しい。その場合は同時にスマートフォンで写真を撮影し、画像文字認識を併せて用いることで精度を上げることができると考えられる。

評価実験では、音のフィードバックが発話に重なってしまい煩わしいという意見があった。また、整合・不整合のフィードバックを待たずに連続して複数項目を入力したいとの意見もあった。このことから、複数項目を連続で入力できるようにするとともに、連続して発話中は音声フィードバックをせず、入力が終わった時点でまとめて音声フィードバックを返すと良いと思われる。その場合には音声フ

ィードバックが長くなるとともにどの項目についてフィードバックを返しているのかがわからないため、フィードバックする音声の内容については注意深く吟味する必要がある。

## 6. おわりに

本稿では、トリガ検出とコマンド型音声認識を用いることで、消費電力やネットワーク環境といった条件を満たしつつ、作業のデジタル化と、作業者が好きな順番・タイミングでの項目入力を実現する作業支援システムを提案した。この項目入力の仕組み自体は汎用的であり、様々な作業に適用することができる。

また、実用性の検証を通して、工数削減にはならないものの、簡単な入力内容であれば実用性があることがわかった。しかし、認識結果や照合結果のフィードバック方法については課題があることもわかったため、フィードバック方法についての改良検討を行った。今後はこれを含めたユーザインタフェースの改良を行っていく。

本稿では簡単な模擬実験のみ行ったが、今後は作業者が慣れた順番・タイミングで入力が行えるか、実際に運用可能かを確かめるため、現場作業者による運用を通じた検証が必要である。また運用に際しては、読み方のルールや照合の曖昧性の範囲等、細かい運用ルールの策定も必要となる。

## 参考文献

- [1] “VOHMIK”. <https://www.asahi-kasei.co.jp/vmk/index.html>, (参照 2018-06-27).
- [2] “Animos Voice@Work”, [https://www.animo.co.jp/for\\_biz/vaw/](https://www.animo.co.jp/for_biz/vaw/), (参照 2018-06-27)
- [3] 田淵仁浩, 坂口基彦, 服部浩明, 奥村明俊. 現場作業支援ソリューションのための音声対話型 AI 帳票. 情報処理学会論文誌コンシューマ・デバイス&システム, 2018, vol.8, no.2, p.13-23.
- [4] “RECAIUS フィールドボイス”, <https://www.toshiba-sol.co.jp/pro/recaius/lineup/fieldvoice.html>, (参照 2018-06-27).
- [5] “Google Home”. [https://store.google.com/jp/product/google\\_home](https://store.google.com/jp/product/google_home), (参照 2018-06-27).
- [6] “Amazon Echo”. <https://www.amazon.co.jp/dp/B071LMG8BQ/>, (参照 2018-06-27).