





# つくば公共サービス共創事業

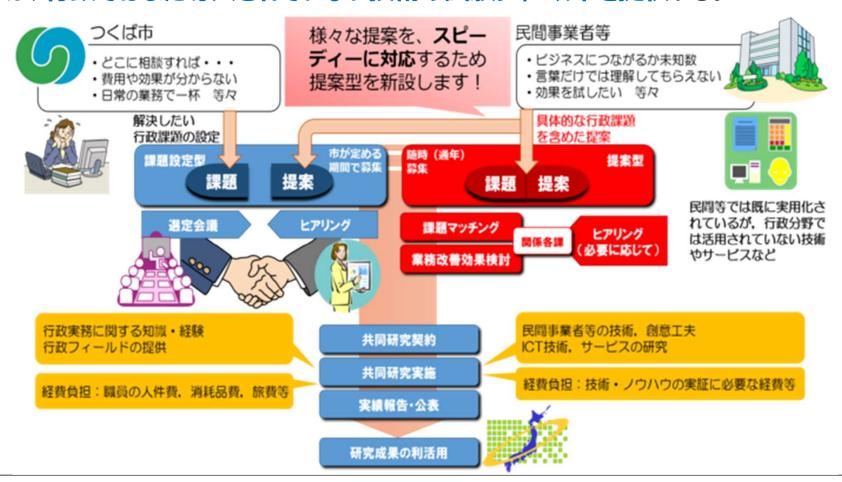
AI技術を活用した議事録等の文字起こし作業自働化による自治体業務の負荷軽減及び効果

つくば市 株式会社NTTデータ



### つくば公共サービス共創事業とは

- つくば公共サービス共創事業(通称、「つくばイノベーションスイッチ」)とは、既に民間では導入が進んでいる先端ICT技術が、まだ公共サービスの分野では導入が進んでいないことに着目して、それらの先端ICT技術を、市民サービスの向上及び行政課題の解決等に資する新しい製品やサービスの創出に結びつけるため、民間事業者等との共同研究を実施するものである。
- つくば市は、行政ではまだ導入されていない技術の試験フィールドを提供する。



### 2019年度の共創事業への提案に関して

■ 2019年度のつくば公共サービス共創事業へは、「AI技術を活用した議事録等の文字起こし作業自動化による自治体業務の負荷軽減及び効果」を共同研究対象テーマとしてつくば市は提案を受けた。

#### 解決する行政課題

自治体業務における情報公開や透明性確保は、自治体経営において重要な課題の一つであり、そのための情報発信手段の一つである議事録等の作成業務は、SNSやデジタルデバイスの普及に伴い、更なるスピード化や正確性が求められる状況であり、自治体職員の業務負荷となっている。

具体例1

庁内で開催される公開/非公開会議の議事録等 を**職員や外部委託先が手作業で文字起こし**する。

具体例2

災害発生時等の緊急時、**文字起こしや確認、修 正作業に多くの時間が発生**している。

具体例3

外部委託先への文字起こし委託に伴い、**発注から納品までに多くの費用と時間が発生**している。

#### 共同研究による解決を考えている課題

- ✓ AI技術を活用した文字起こし作業の自動化による作業時間の短縮(効率性向上)
- ✓ AI技術を活用した文字起こし作業の自動化による正確性の向上 (生産性向上)
- ✓ AI技術を活用した文字起こし作 業の自動化による費用低減(費 用対効果向上)

### なぜAI技術を採用するのか?

- 生産性向上の必要性の高まり、人工知能(AI)技術の革新の両面から、人工知能(AI)を活用した業務の効率化が脚光を浴びている。
- 高度なAI技術を用いたシステム導入に着手する前に、AI技術を一部取り入れ、迅速に導入が可能か つ即効性を得ることができるAIソリューションが注目されている。

#### 生産性向上の必要性

生産年齢人口の減少

働き方改革

AI技術を活用した 業務の効率化

> AI ソリューション

#### AI技術の革新

ディープラーニングによる 認知能力の劇的向上

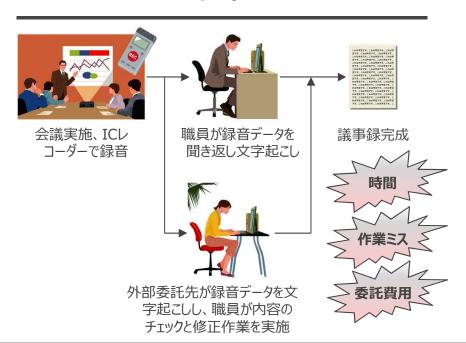
複雑化する大規模システム (システム全面刷新の難易度高)

### 共同研究の目的

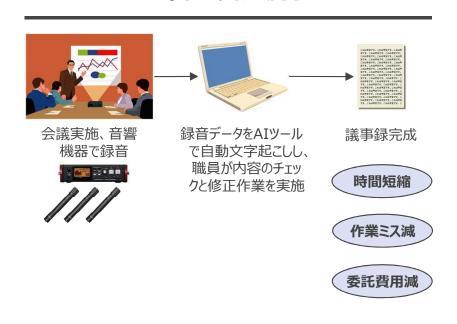
- つくば市では、各種会議の議事録を作成するために、職員がICレコーダーの録音データを何度も聞き返しながら作業を行っており、会議時間の何倍もの時間を要している。
- また、会議終了から議事録作成までの期間短縮の要望もあり、議事録作成に対する職員の業務負担が課題になっている。
- 今回の共同研究により、AIソリューションによる議事録の文字起こしを利用した場合にどの程度の業務 負荷軽減につながるかどうかを検証することを目的とする。

#### 【行政課題の解決後のイメージ】

#### 従来



#### 共同研究実施後



### 共同研究者の選定

- 2019年9月に株式会社NTTデータより共同研究に関する共同研究申請があった。
- 選定事項の確認の結果、前述事業者を共同研究者として選定し、10月29日に契約を締結した。

選定日程	選定実施内容
2019年9月3日	共同研究者より共同研究申請書を申請
2019年9月30日	選定結果通知(NTTデータを共同研究者として選定)
2019年10月29日	契約締結、共同研究開始

### 共同研究の進め方

■ 本共同研究は、下記の手順で実施した。

共同研究キックオフ 2019/10/8 本共同研究の目的、進め方、ゴールイメージの共有、課題の洗い出しをメンバー全員で実施する。

環境準備、トレーニング 2019/10/17 AIソリューションを利用するための端末環境を準備し、使用機器の操作方法を含めたトレーニングを実施する。

対象会議選定開始 2019/10/18

庁内でのAIソリューションによる議事録作成業務の自動化に対する希望調査を実施し、共同研究の対象会議を選定する。

共同研究開始 2019/10/31

実際の会議の模様を録音し、AIソリューションによる文字起こし及びAIソリューションの言語学習を実施する。

検証データ分析作業開始 2020/1/24

会議の録音データ及び文字起こし後のテキストデータに基づき、共同研究の効果測定に必要な検証データを分析する。

中間報告実施2020/2/18

検証データ分析状況を踏まえ、最終報告書作成に向けた整理の方向性、残課題の洗い出しを実施する。

最終報告書作成開始 2020/3/2

共同研究の成果として、AIソリューションの導入効果、今後の展開について、最終報告書としてまとめる。

# 共同研究について

### 共同研究の概要と研究手順

共同研究の概要

クラウドサービス型のAI議事録ソリューション「AI Minutes for Enterprise」を利用し、つくば市の議事録等の作成業務を自動化する ことで、業務の負荷軽減効果や付加価値を測定し、導入に向けての課題抽出を実施する。

■ 利用したAIソリューション クラウドサービス型AI議事録ソリューション「AI Minutes for Enterprise」(以下「AIM」という)

#### 共同研究手順

- 1. 事前調整
- ①議事録起こしに手間がかかっている会議があるか庁内アンケートを行い、導入による効果検証対象会議の洗い出しを実施
- ②AIMの庁内向けデモを実施



果が高そうな会議を洗い出し





AIMの庁内デモ実施による 庁内の意識醸成

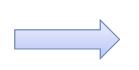
#### 2. 個別会議

- ①会議の音声データを録音
- ②AIMに音声データをアップロードし、テキストへの書き起こしを実施
- ③書き起こし結果を確認し、適宜テキストの修正を実施
  - ※過去の議事録等をAIMに読み込ませ、学習させることで、音声認識率の向上による作業負担の軽減効果を狙った
- ④これまでの業務とAI議事録ソリューションを利用したことでの作業時間等の比較、検証を実施



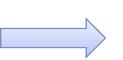


自動書き起こし





自動書き起こし結果の 確認、テキストの修正





作業時間等の比較、 検証の実施

### 共同研究で使用した機器

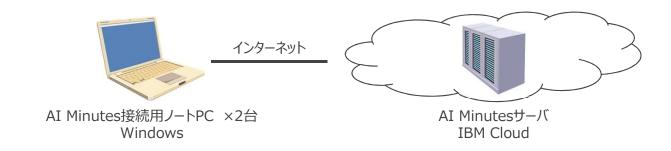
■ 会議音声録音



■ 音声書き起こし



■ 書き起こし結果確認・修正・テキスト取得、認識モデルカスタマイズ(学習)



## AIソリューション適用の検証結果1

•議事録作成業務

・テキスト入手までの待ち時間

・文字認識精度の検証について

### 議事録作成業務の削減

■ 議事録の形式(逐語録・概要まとめ)の違い、発話者の話し方などによってばらつきが見られるものの、逐語録の場合、25%~60%程度の削減効果が確認できた。一方、概要形式の場合では、約70%の削減効果が確認できた。

会議名	議事録形式	議事録作成時間	適用後作成時間	削減効果
会議A		4時間	3時間	1時間 (25%)
会議B		12時間	9時間	3時間 (25%)
会議C	逐語形式	7時間	4時間	3時間 (43%)
会議D		5時間	2時間	3時間 (60%)
会議E		8時間	5時間	3時間 (38%)
会議F		3.5時間	2時間	1.5時間 (43%)
会議G		21時間	10時間	11時間 (52%)

会議名	議事録形式	議事録作成時間	適用後作成時間	削減効果
会議H	Interest In	0.83時間	0.25時間	35分 (70%)
会議I	概要形式	2時間	0.5時間	1.5時間 (75%)

※議事録起こしを外部委託せず、職員自身が文字起こし作業を行っている会議を対象とした。

### テキスト入手までの待ち時間の削減

- 外部委託先に委託する場合、議事録を受け取ることができるまでに数日を要することが一般的だが、 本ソリューションを活用すると数時間で議事録を完成することができた。
- 全ての会議でテキスト入手までの待ち時間は大幅に削減できた。

会議名	会議名     適用前		削減効果
会議(イ)	336時間(14日)	10時間	326時間
会議 (口)	240時間(10日)	8時間	232時間
会議(八)	240時間(10日)	4時間	236時間
会議(二)	240時間(10日)	22.35時間	217.65時間

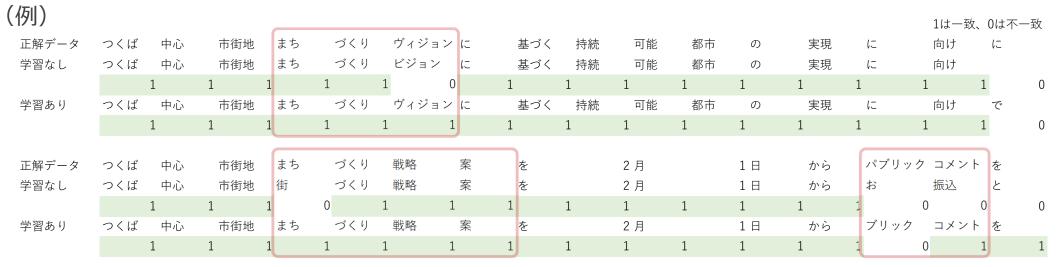
### 文字認識精度の検証について

#### 【検証事項】

- AIエンジンへの事前学習の有無による正解率の変化
- 話者や発言内容の違いによる正解率の変化

#### ● 認識精度計算方法

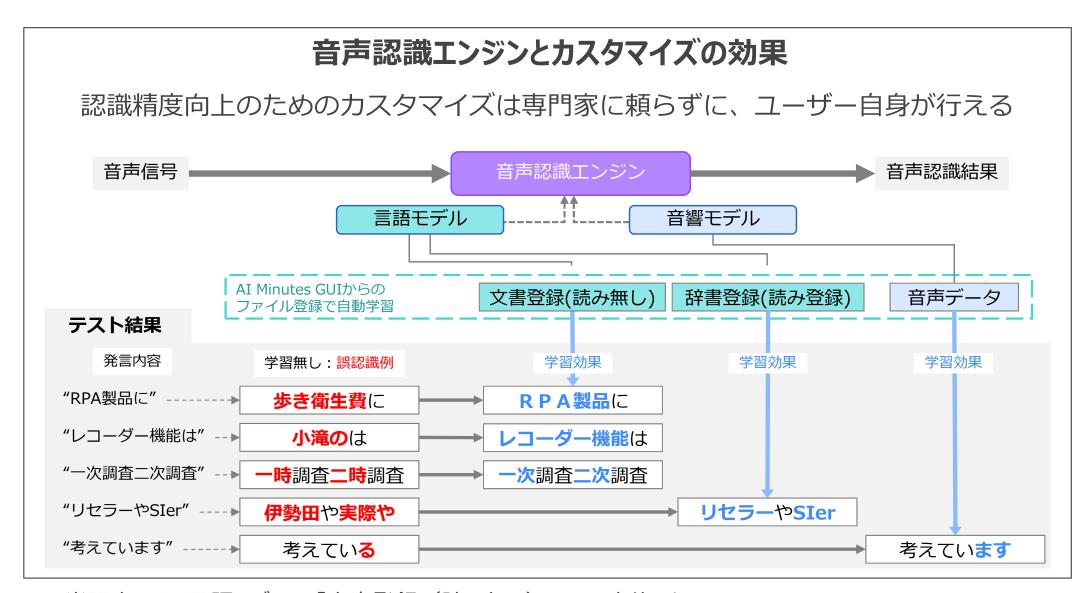
各条件下で音声認識されたテキストを形態素(言語上の最小単位)に分割し、正解データに対しそれぞれが一致しているか否かを集計し、正解率を算出する。



#### 【検証結果】

- 事前学習をさせることによって、正解率が向上する傾向が見られた。
- 話者や発言内容により、正解率に違いが見られた。

### (参考) 事前学習(カスタマイズ)による効果検証



■ 当研究では、言語モデル - 「文章登録(読み無し)」のみ実施した。

### 文字認識正解率について(教育関連1)

#### 【会議a】

	正解率(正解数÷正解データ単語数)(%)				
	学習あり(1)	学習無し(2)	正解率差(1)-(2)		
話者A	91.7	90.3	1.4		
話者B	83.9	81.9	2.0		
話者C	70.8	68.1	2.7		

	全語	吾数	正角	<b>挥数</b>	正解データ単語数
	学習あり	学習なし	学習あり	学習なし	止所ナーグ半品数
話者A	153	155	132	130	144
話者B	168	167	125	122	149
話者C	166	166	102	98	144

- 話者によっては、文字認識の正解率の差が20ポイント以上見られた。各話者の発言内容に偏り(専門用語や地域特有の固有名詞などが含まれる量の違い)はなかったため、マイクとの距離や話し方の違いによる影響が大きいのではないかと考える。
- 各話者においても、学習による正解率向上効果が、1~3ポイント程度だが見られる結果となった。

### 文字認識正解率について(教育関連2)

#### 【会議b】

	正解率(正解数÷正解データ単語数)(%)					
	学習あり(1) 学習無し(2) 正解率差(1)-(2)					
話者A	97.7	97.7	0			
話者B	92.0	92.0	0			

	全語	全語数		正解数	
	学習あり	学習なし	学習あり	学習なし	正解データ単語数
話者A	136	135	128	128	131
話者B	167	165	149	149	162

- 2話者とも90%以上の正解率となった。特に話者Aの発言は正解率が高かった。この発言は、冒頭の他の委員に対する挨拶であったため、通常の発言以上に、話す速度や話し方に注意が向けられていたのではないか。
- 発言内容に専門用語や地域特有の単語などはあまり含まれていなかったため、学習による効果は見られなかった。

### 文字認識正解率について(教育関連3)

#### 【会議c】

	正解率	(正解数÷正解データ単語数)	(%)
	学習あり(1)	学習無し(2)	正解率差(1)-(2)
話者A	89.6	85.6	4.0
話者B	80.8	80.8	0

	全語数		正解数		正规二 万半宝米
	学習あり	学習なし	学習あり	学習なし	正解データ単語数
話者A	124	125	112	107	125
話者B	137	136	105	105	130

- 話者の違いによって正解率に約5~9ポイントほど差が見られた。話者Bの発言は、議論の流れの中の発言だったのに対し、話者Aの発言は単独で完結する儀礼的、形式的な発言だったため、通常の発言以上に、話す速度や話し方に注意が向けられることによって、正解率がより高くあらわれたのではないかと思われる。
- 発言内容に専門用語や地域特有の単語などはあまり含まれていなかったものの、学習による効果が正解率の差となって若干見られた。

### 文字認識正解率について(その他分野1)

#### 【会議d】

	正解	率(正解数÷正解データ単語数)	(%)
	学習あり(1)	学習無し(2)	認識率差(1)-(2)
話者A	86.2	81.4	4.8
話者B	91.5	87.2	4.3
話者C	95.6	92.6	3.0

	全記	吾数	正解数		正解データ単語数
	学習あり	学習なし	学習あり	学習なし	止所ナーツ半品数
話者A	172	165	125	118	145
話者B	159	158	129	123	141
話者C	158	154	130	126	136

- 話者の違いによって正解率に約4~9ポイントほど差が見られた。今回の話者A~Cの発言は、議論の流れの中の発言だった。他会議との単純な比較はできないが、文章の読み上げや儀礼的・形式的な発言に比べ、その場の状況に応じて発言するケースは正解率が低下する傾向が見られる。
- 発言内容に専門用語や地域特有の単語などはあまり含まれていなかったものの、それらの単語以外の修正が入っており、学習による効果が約3~5ポイント見られた。

### 文字認識正解率について(その他分野2)

#### 【会議e】

	正解率(正解数÷正解データ単語数)(%)				
	学習あり(1)	学習無し(2)	正解率差(1)-(2)		
話者A	100	97.3	2.7		
話者B	97.7	95.3	2.4		
話者C	99.2	92.2	7.0		
話者D	100	97.9	2.1		
話者E	99.3	97.8	1.5		

	全語数		正解数		TCA77" 万比克米
	学習あり	学習なし	学習あり	学習なし	正解データ単語数
話者A	147	148	147	143	147
話者B	129	128	126	123	129
話者C	129	127	128	119	129
話者D	143	143	143	140	143
話者E	141	139	137	135	138

- 今回の話者の発言は、ほとんどが正解率95%以上という結果であった。今回の話者A~Eの発言は、市側の事務局の発言であったり、挨拶的内容の発言を選択して、率の算出を行ったが、他会議同様、やはり文章の読み上げや儀礼的・形式的な発言に加え、相手に対する説明や質問に対する回答は、相手に内容を伝えようと発言速度や話し方に注意がより傾き、正解率の差としてあらわれた可能性がないか。
- 発言内容に専門用語や地域特有の単語などはあまり含まれていなかったものの、それらの単語以外の修正が入っており、学習による効果が約2~7ポイント見られた。

# AIソリューション適用の検証結果2

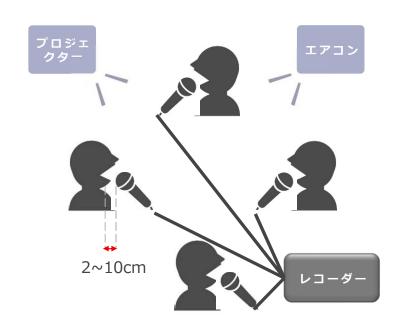
・全指向性マイク利用検証

・オンライン会議システムの利用検証

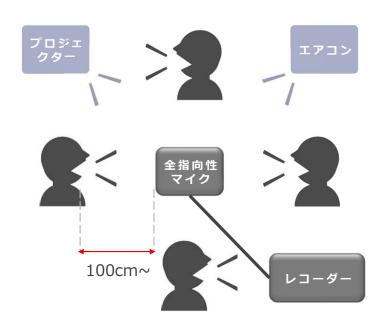
### 全指向性マイク利用検証

#### 全指向性マイクを利用した実証を行った背景、課題は何か?

- 今回の共同研究では、音声品質の高い集音を実現するために、主に感度の高い**単一指向性のマイクを 利用して録音**を行ったが、会議・打ち合わせの内容によってはより簡易的な形式のものもあるため、より気軽 に録音ができるよう全指向性のマイクをセッティングするなどできると利用の幅が広がるように感じる。
- ただし、録音時の音質をどれだけ担保できるかの問題もあるため、全指向性のマイクについては、**一定の音質を確保できる製品を選定**する必要がある。
- 単一指向性マイクの場合、機材の点数は多めになり、セッティングに労力がかかる場合がある。一方、全指向性マイクはレコーダー機能を兼ねた機材もあるため、機器運用の観点では容易に利用できる。



単一指向性マイクでの集音



全指向性マイクでの集音

### 全指向性マイク利用検証

- 全指向性マイクを備えたレコーダーを使用して音声収録を行いテキストへの書き起こしを試みたが、音声認識がうまくされず会議音声書き起こしに活用できる水準ではなかった。
- 収録音声を聞いてみたところ、話者の声は**遠くで話しているような音**となっており、加えて会場で発生する様々な音もかなり混ざっていた。



TASCAM DR-100mkⅢ

### 全指向性マイク利用検証:検証結果の考察

- 全指向性マイクを使用した環境について音声認識の観点では、以下のような人間の声以外の音が 認識精度を下げる要因となるため、これらに配慮した収録方法を総合的に検討する必要がある。
  - 部屋の反響音:音の要素が短い時間差で2重3重に重なる
  - 各種ノイズ:設置機器(空調・プロジェクタ等)が発する音、机上の物品(ノート、ペン、カップ等)が机に当たる音
  - 人間の動作音: 部屋の出入り、咳払い、小声の会話

### オンライン会議システムでの利用検証

オンライン会議を利用した実証を行った背景、課題は何か?

- 新型コロナウイルスに関連した緊急時における当該ソリューションの効果
  - 今回の新型コロナウイルスの影響下においては感染拡大防止の観点から、人と人との接触を可能な限り減らすことが重要と考えられている。
  - 本市においても本ウイルス対策を検討するにあたっては、従来の人が実際に会議室に参集して 議論をするというやり方から、オンライン会議システムでの議論を行うことが基本となった。
  - 会議のあり方の潮流がこれまでから変化しているものの、議事録作成の重要性が変わることはなく、むしろ刻一刻と変化する情勢に対し適切に判断を行っていくためには、過去にどのようなことを議論したかといった内容が、より速いスピード感で求められている。





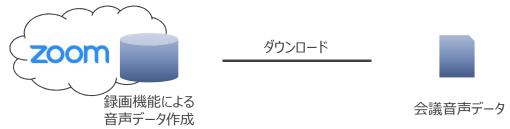






### オンライン会議システムでの利用検証:使用した機器

■ ツール内蔵録音機能:会議音声録音



■ ツール内蔵録音機能:音声書き起こし



■ ポータブルレコーダー録音:会議音声録音



■ ポータブルレコーダー録音:音声書き起こし



### オンライン会議システムでの利用検証:文字認識正解率

#### 【検証事項】

- 音声録音方法の違いによる認識率の変化
- 話者の違いによる認識率の変化

#### 【検証結果】

- ツールに内蔵されている録音機能を使用した方が、正解率が向上する傾向が見られた。
- 話者の違いによる正解率には違いが見られた。

	正解率(正解数÷正解データ単語数)(%)					
	ツール内蔵機能録音(1)	ポータブルレコーダー録音(2)	認識率差(1)-(2)			
話者A	93.0	73.0	20			
話者B	80.0	69.5	10.5			
話者C	91.6	79.0	12.6			
話者D	86.7	88.0	-1.3			
話者E	88.6	81.8	6.8			
話者F	83.7	81.5	2.2			

	全語数		正解数		正解データ単語数
	内蔵録音	ポータブル	内蔵録音	ポータブル	止所ナーツ半詰奴
話者A	104	98	93	73	100
話者B	91	92	76	66	95
話者C	124	126	109	94	119
話者D	78	78	72	73	83
話者E	86	84	78	72	88
話者F	91	93	77	75	92

### オンライン会議システムでの利用検証:職員の感想

- これまでの利用方法では話者がマイクを介して話していたが、今回のオンライン会議では大多数が、マイクを使わずに (※1) 話をしていたため、音声認識ができる品質の録音ができているか不安であったが、通常の利用方法と大差ない認識率であったので、効率的に文字起こしをすることができた。 (※1) 遠隔参加者はスマートフォンなどで参加し組み込みマイクかマイク付きイヤホンを使用、庁議室ではハンドマイクを使用
- 実験件数がまだ少ないため、何とも言えないが、ツール内録音機能で今回 のような結果が今後も得られるようであれば、録音機器などにコストを費やす ことが少なくなるため、より安価に導入できる可能性もあるのでは。

### オンライン会議システムでの利用検証:検証結果と利用者感想 に対する考察

- Web会議という性質上、以下の要素も精度が上がる要素
  - マイク(PCマイク/マイク付きイヤホン)に向かって発言するという意識が働き、ノイズの少ない良好な音声収録が期待できる。
    利用者感想「マイクを使わずに話をしていた」という状態は、実際にハンドマイクは使わなかったがWeb会議のマイクに向けて意識して発言していたものと解釈できる。
  - 発言順序がある程度制御されるため、音声が重なり認識結果に影響することの低減が期待できる。
- Web会議システムや使用デバイスの特性のメリット
  - Web会議システムには背景雑音の抑制機能があり、ノイズの低減が期待できる。音量も自動調整される。
  - スマートフォンやマイク付きイヤホンは発言者の近接にあることが多くなるため、音質が良くなることが期待できる。
- 録音機能の利便性
  - 録音機能で得られた音声データの方が概ね音声認識精度が高くなることが確認できた。レコーダー録音では録音までの経路で音質劣化の要素があるためと解釈できる。
  - 当初の利用形態が録音をベースにしている(リアルタイム書き起こしでない)場合、ツールの録音機能は録音機材の代替になりうる。
  - 録音データは多くの場合クラウド上に置かれるため、保管領域の容量不足の心配は低減できる。



### 共同研究を担当した職員の感想1

- 録音して、すぐにテキスト化できるので**急ぎで議事録作成が必要な際には助かった。**
- 聞きたい音声を繰り返しピンポイントで聞けるので、**修正作業が楽**だった。(※1)
- 前回分を学習しているため、**特徴的な専門用語もしっかり認識できていて助かった。**
- AIへの事前学習について、**以前の議事録データをコピーアンドペーストするのみだったので、** 非常に簡単であった。(※1)
- 発言者によっては、マイクの使用が不慣れなこともあり、声を拾えていない所もあった。
- (※1)本研究に使用したAI議事録ソリューション(AI Minutes for Enterprise)の機能・特長への評価であり、 AI議事録ソリューション全般に該当するものではない。

### 共同研究を担当した職員の感想2(課題や要望、対応案)

- ■「人が一から音声を聞きながら書き起こすよりは**大幅に負担軽減にはなる**」との意見が多く挙げられたが、**音声認識の精度が不十分な部分も散見**され、当該部分の修正作業が意外に時間を要しているようだった。 (対応案) 音声認識は100%にはならないため、修正作業は必要。今回のAIMのようなツールによっては複数人で同時に修正作業を行えるので、作業を分担することで負荷の軽減や時間の短縮が可能になる。
- 学習量を増やすことで**音声認識の精度をより高める**ことができれば、**修正作業の負担も軽減されると想定**される。

(対応案) 学習量を増やすことでさらなる認識精度の向上は期待できる。さらに、誤変換の傾向を分析し、 その結果に沿った学習を行うことも効果的。

■ 人間が聞き取りづらい言葉は、無論AIも認識が困難になる傾向が強い。より音声認識の精度を高めるためには、学習量を増やすとともに、マイクを正しく使い滑舌良く発話する等、話し方にも注意を払ってもらう必要がある。

(対応案)マイクの向きや距離の固定、滑舌の良い話し方などをガイドし、それを実践してもらうことで精度向上が期待できる。

■ 出力形式と議事録として報告する形式との間に違いがありすぎて、報告形式に変換されるまでの手間が大きいように感じられた。出力形式については、ある程度一般的な形式で出力されるとよかった。 (対応案) 報告形式が定型なのであれば、出力されたテキストから報告書に移す作業をマニュアル化することも手段の一つと考えられる。またRPAなど将来の自動化ツール開発なども検討の候補になる。



### 共同研究者の総括

- 自治体の業務は広範にわたることもあり、関係者の広さ・交わりも多く、会議は非常に多く存在する中で、自治体の情報公開や透明性確保のための情報発信手段の一つである議事録等の作成業務は、重要な位置づけをしめる一方、どの自治体においても置かれている状況は類似しており、業務負担となっていると思われる。
- こうした背景の中、今回の研究結果により、AI議事録ソリューションが議事録等の音声文字起こし業務の負担軽減に寄与することが確認できた。
  - 一言一句を文字化する「逐語録形式」の場合は3割程度、全体の大まかな内容をまとめる「概要形式」に関しては、7割ほどの削減効果を確認できた。
  - 外部委託で受け取りまでに10~14日必要だった議事録も4~23時間の作業でテキストを入手でき待ち時間を大幅に削減できた。
- 検証した『AI Minutes for Enterprise』の大きな特徴として、「学習機能」があった。過去の議事録テキストをコピー&ペーストで読み込ませるだけで、システムが学習できるのは非常に簡単で使いやすかった。庁内で簡単にカスタマイズ化することが可能で、音声認識率のさらなる向上が期待できる。この機能を活用し、音声認識正解率が2~7ポイント向上し100%を達成した書き起こしもあった。このようにAIエンジンの能力だけでなく、使用者が認識精度を日々高めていくことも効率化促進のためには重要である。
- 今後、技術革新の展開により音声認識の精度がより向上すれば、音声文字起こしの業務のより一層の効率化はもとより、音声認識を絡めたサービスへの波及によってさまざまな自治体業務の負荷軽減も期待されると思われる。
- 多様化する市民ニーズに伴い、AIの力をより活用することで、職員が人でなければ行うことが難しい創造的業務に注力できることで、よりよい行政サービスの提供にAIの力を活用していければよい。

つくば公共サービス共創事業 ~つくばイノベーションスイッチ~ AI技術を活用した議事録等の文字起こし作業自動化による自治体業務の負荷軽減及び効果 共同研究実績報告書

> 2020年10月第1版発行 Copyright © 2020 City of TSUKUBA, NTT DATA Corporation 複製厳禁·無断転載禁止