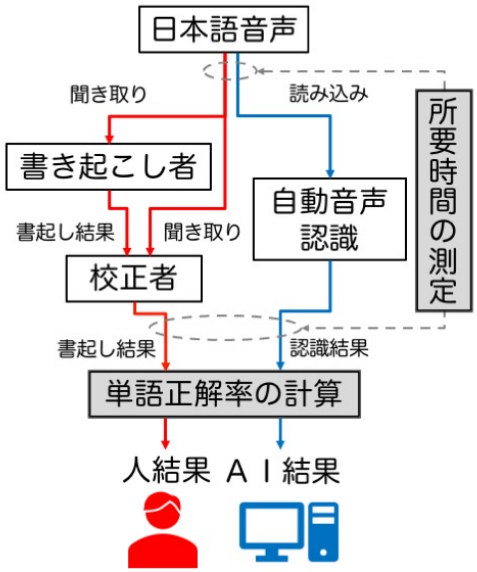
A05

自動音声認識と人との文字起こし能力比較

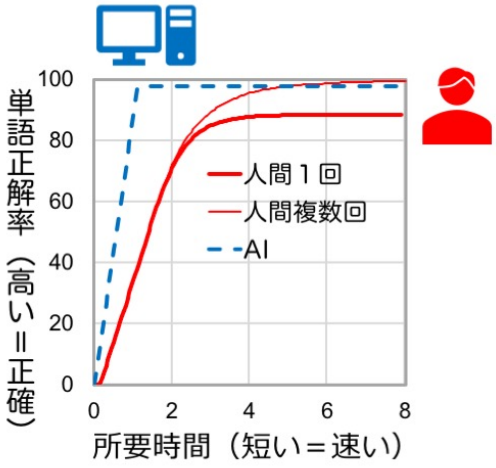
～ AIは人間を超えたか？ ～

概要

スピード勝負では既に大きく人を凌駕する自動音声認識ですが、あらゆる面で人間に勝ったわけではありません。現時点でのAI技術の限界がどこに現れるのかを解説します。



人 vs. ＡＩ競争の流れ



人 vs. ＡＩ結果

特徴

- NICTの多言語音声翻訳アプリVoiceTra®の実力を紹介します
- 人間と“ガチで”競わせてみました
- 現状の到達点を把握し次のステップを探ります

過程と結果

- 日本語音声を聞いて文字にする能力を人間側代表の書き起こし熟練者と厳密に対等な条件で競わせました
- 速さではAIが圧勝、正確さでは…
  1度しか聞けない場合はAIに、聞き直しできる場合は人間に軍配が上がりました

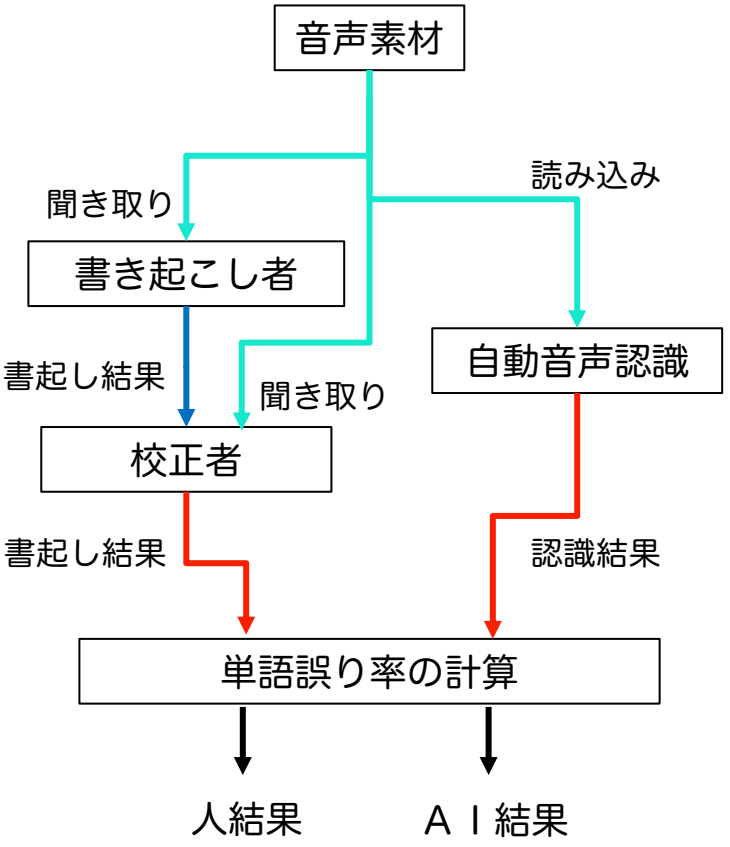
今後の展開

- ＡＩが人間に至らない点は文字に直接現れない手がかりの利用能力に現れます
- 韻律や意味などから人間なら常識的に間違わない部分です
- ＡＩに常識を植え付ける研究に期待がかかります

【お問合せ先】
国立研究開発法人情報通信研究機構 ユニバーサルコミュニケーション研究所
先進的先進的音声翻訳研究開発推進センター 先進的音声技術研究室
Mail: astl-service@ml.nict.go.jp

人 vs AI自動音声認識比較の概要

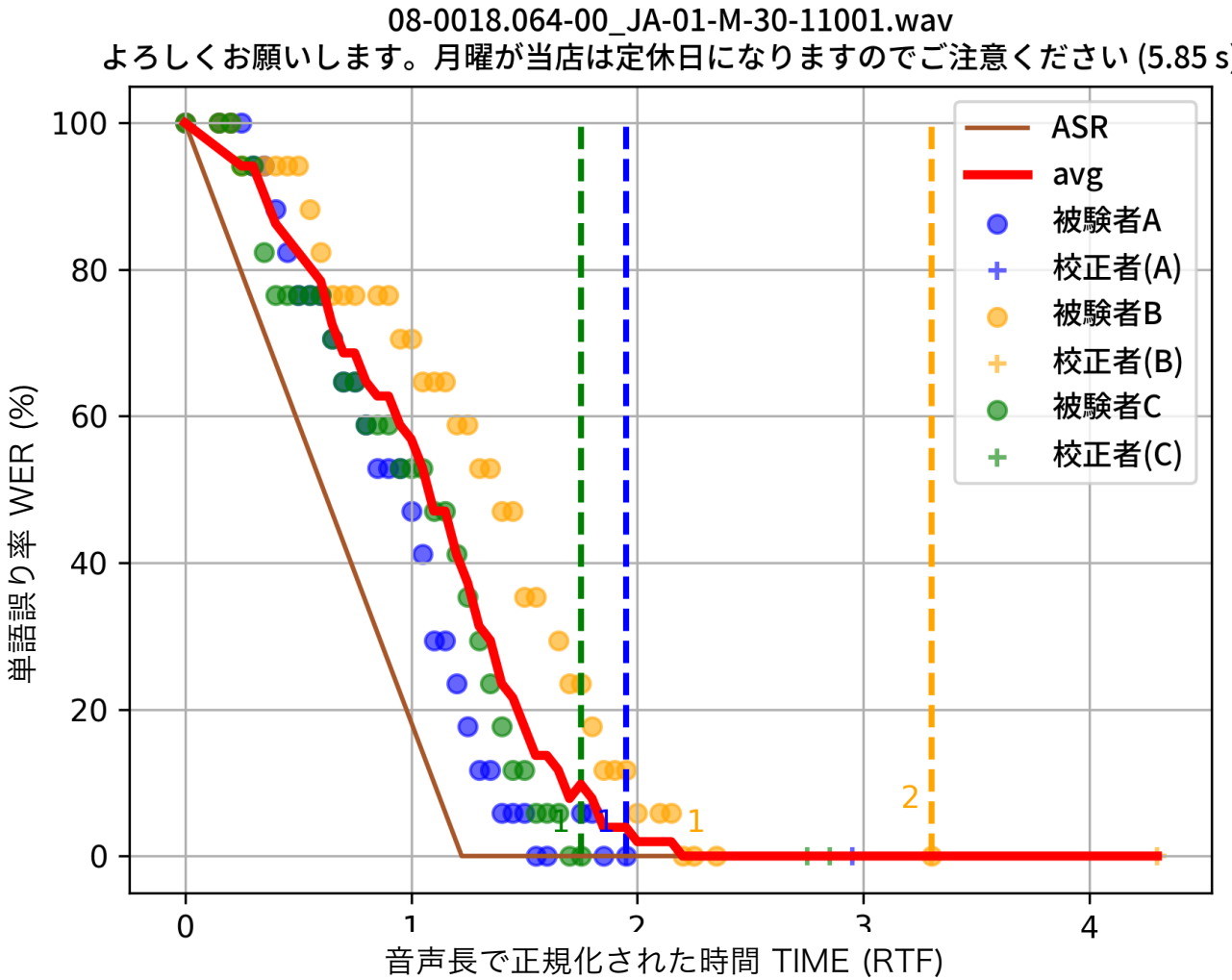
- 参加者
  人：プロの日本語音声書起し者(3名)と校正者（熟練者）
  AI：VoiceTra®（商用APIサービスと同等性能）
- 音声素材
  20名のプロではない成人男女が発話した日本語700文・700音声（文の重複なし）
- 評価指標
  正確さ：単語誤り率 WER(=[置換＋脱落＋挿入]／単語数)
  速さ：正規化所要時間 (=実所要時間／音声時間長 (RTF))
- 手順
  - 書き起こし者はヘッドホンで各音声を１回聞き取り、PCのキーボードにより内容を可能な限り書き起こした。
  - 必要ならば２回以上聞き取り、間違いがないと判断した時点で終了した。3名とも700音声を書き起こした。
  - 校正者は同じ方法で書き起し結果を修正した。
  - ＡＩは音声ファイルを読み込み自動音声認識を行なった。
  - ＡＩの所要時間は音声が実際に再生された場合を想定し、再生開始時点を起点として認識結果が出るまでの経過時間とした。
  - 人の所要時間は専用のアプリを作成して精密に測定した。



人 vs.ＡＩ文字起こし実験の流れ

専用書き起こし制御・測定ツール

典型例

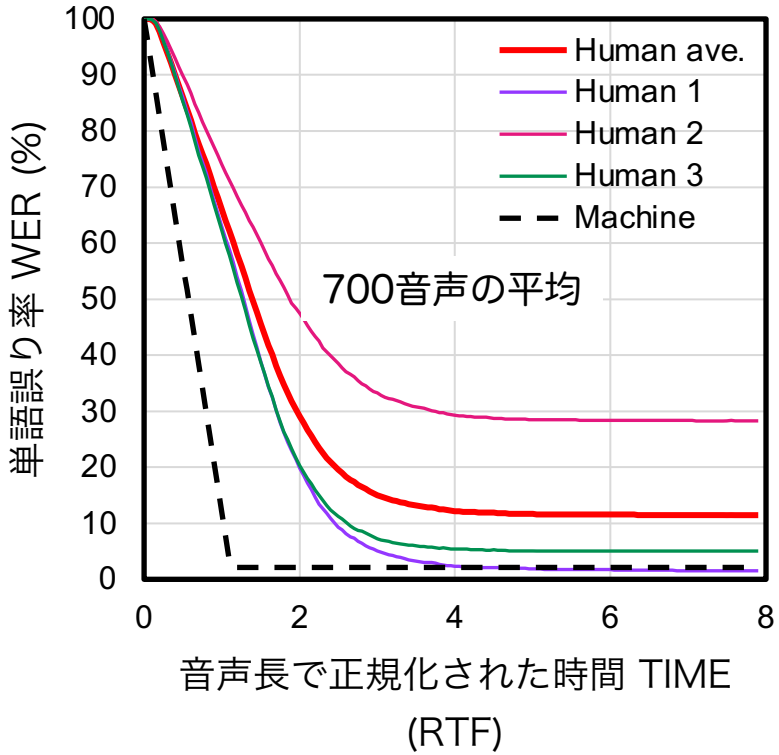


典型例

AI (自動音声認識)より人(avg)が少し遅れて誤り率0に到達する。
("1","2"の破線は1回目、2回目の再生による書起しが終了した時点)

1回の再生のみで書き起した場合：AIは人の文字起こし能力を超えた

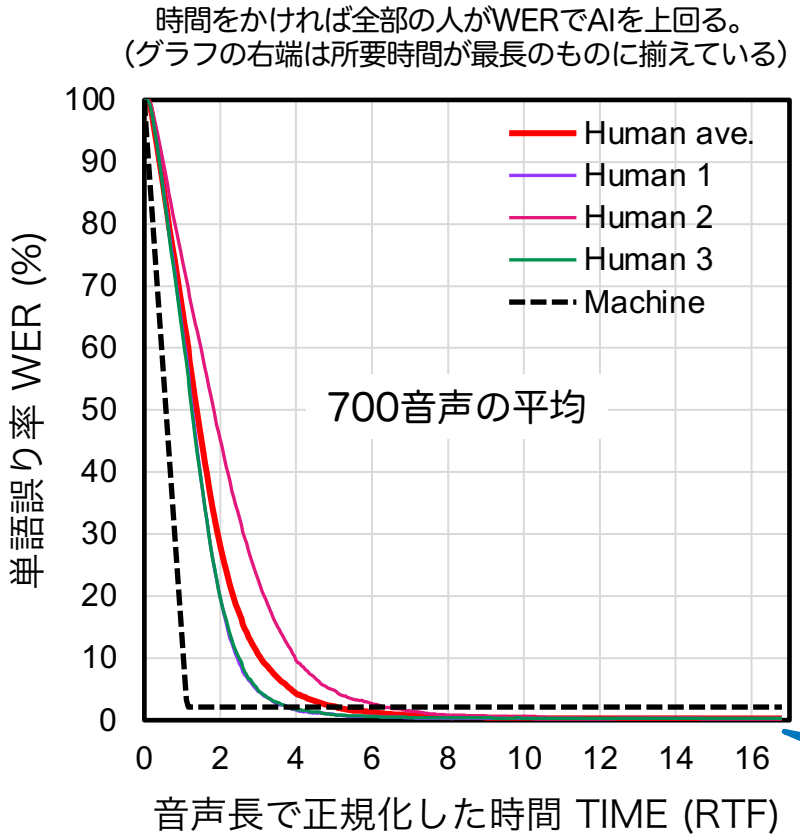
1名（Human 1）を除き、人のWERはAIに及ばない。
（グラフの右端は所要時間が最長のものに揃えている）



	Humans	AI	
WER (%)	11.45	2.12	（正確さ）
Time (RTF)	2.26	1.11	（速さ）

- 公平な比較条件
- 速さはAIが人の２倍以上
- 正確さでもAIが人を大きく上回った

複数回の再生を許した場合：
時間をかければ正確さは人がAIに勝る



	Humans	AI	
WER (%)	0.24	2.12	（正確さ）
Time (RTF)	4.42	1.11	（速さ）
#playback	2.46	1	

厳密には公平ではないが、従来報告で採用された条件
最終的には全ての書き起こし者が正確さでAIを上回った

AIが人を超えられなかった部分

AIが人を超えられなかった例

	人の書き起こし (correct)	ＡＩの音声認識 (incorrect)	正誤の理由
①	送信者にユーザーが含まれている場合，オンにします。	～本にします。	意味の整合性
②	カニですね，どの辺が，カニに見えるのですか？	～仮に見える～	意味の整合性
③	まずは型名を教えてください。	まずは片目を～	韻律要素
④	根津には，東京メトロという地下鉄を利用して行きます。	ネズミは，～	韻律要素
⑤	プロテスタントでは，牧師と言うのですが，たぶんそうすると思います。	～ボックスと言うのですが～	意味の整合性，韻律要素

★意味の整合性や音韻長・アクセントなど韻律要素の利用では人が優位性を保つ

まとめ

人とAIの自動音声認識との文字起こし能力を比較

1回だけの再生では、AIは人の能力を速さ・正確さの両方で超えた（＝厳密に対等な条件）

複数回の再生では、正確さで人が上回った

AIが人に劣る「僅かな」部分：人はまず間違えない

- 開発者側：研究開発への課題
- 利用者側：ＡＩとの付き合い方の知恵