

情報通信研究機構

NICT 先端研究

202

チェスの世界王者を破った人工知能(AI)ディープ・ブルーの登場以来、さまざまな分野で「AIが人間を超えた」との報告が相次いだ。音声入力に使われる自動音声認識技術でも同様の報告が出始

ているが本当にそう
 か？音声認識の現時点
 の到達点を正確に知
 り、さらなる性能向上
 のために何が必要かを
 探るため人間と「ガチ
 で」競わせてみた。
 AI側の代表選手は
 私たちが開発し無償公開
 する多言語音声翻訳ア
 プリ「VoiceTr
 a」である。自動音声

認識部は一般に利用可
 能な商用アプリとおお
 むね同等の精度である
 ことを確認している。
 人間側は日本語音声書
 き起こしの熟練者3人
 と校正者1人の体制と
 した。校正者はAI側
 では原理的に生じ得な
 いタイプミスを修正す
 る役割で加わった。

課題は情報通信研究
 院（NICT）の音
 声認識評価公開デー
 タセット「SPRED
 S2」所収の日本語音
 声700文の文字起こ
 し。速さと正確さの両
 面を競う。速さは所要
 時間の（各音声の持続時
 間長に対する相対時間
 で定義）、正確さは単
 語正解率が指標だ。

まず、人間もAIも
 結果となり、この厳密

音声を1回だけ再生聴
 取しただけで書き起こ
 した場合、所要時間は
 AIの1・1（700
 音声の平均値。以下同
 じ）に対して人間は3
 人の平均が2・3で2
 倍以上かかった。正確
 さにおいてもAI（97
 ・9％）は人間の平均
 値88・5％をしのぐ

音声の平均

時間 (速さ指標)

— 人間1回
 - - 人間複数回
 - . AI

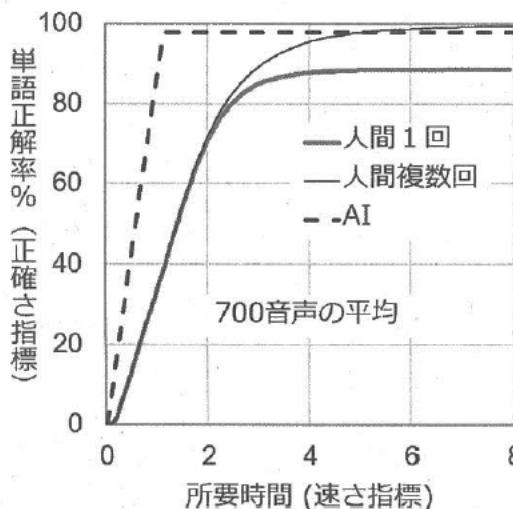
4 6 8

音声認識 AIは人間を超えたか？

ユニバーサルコミュニケーション研究所
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主任研究員

加藤 宏明

99年神戸大院自然科学研究科博士課程修了。国際電気通信基礎技術研究所（ATR）で勤務後、11年より現職。一貫して聴覚と音声を介したコミュニケーションに関わる基礎的研究に従事。博士（工学）。



人間は1回だけの聴取（太実線）では
AI（破線）に速さ・正確さともにか
なわないが、複数回の聴取（細実線）
なら正確さで優れる（NICIT提供）

AIが人間に至らない部分は、「送信者にユーザーが含まれている場合オンにします」という音声で「一本にします」と間違えるなど、音響的に近く文法的にも正しいが意味的にあり得ない例が多い。人間なら常識的にあり得ないと分かる候補を常識のないAIが拾ってしまった結果だ。AIに常識を植え付ける研究に期待がかかる。

(火曜日に掲載)