

音声ファイルを介した言語能力測定可能性の検証

長田 颯斗[†]荒牧 英治^{††}宮部 真衣[†][†]和歌山大学システム工学部^{††}奈良先端科学技術大学院大学研究推進機構

1 はじめに

近年、客観的に言語能力を測る需要が様々な場面で高まっている。例えば、外国人留学生数は近年増加傾向にあり¹、外国人留学生を採用する企業も年々増加している [1]。採用選考における外国人留学生の日本語能力評価に課題を抱えている企業もあり、客観的な言語能力測定が求められている²。また、認知症者も近年増加傾向にある。認知症は、言語能力に何らかの特徴が表れる可能性があることが知られている。Kemper らは、英文において、認知症の進行は構文能力と相関関係にあり、症状の進行度にあわせて構文能力の顕著な低下がみられることを示しており [2]、言語能力を測ることにより、早期発見や療養に役立つ可能性がある [3]。このように、様々な場面で言語能力測定は有用であると考えられる。

これまでに我々は、言語能力測定システム「言秤」を開発してきた [4]。言秤は、テキストや音声発話により入力された内容をもとに、6つの言語能力指標により利用者の言語能力を測ることができる。実験の結果、6つの指標のうち、認識誤りに頑健な Type・Token 割合 (Type Token Ratio; TTR) には発話内容と音声認識結果に強い正の相関があることを示した。ただし、音声による測定においては、マイクでの直接入力を前提としており、より手軽な測定環境が必要である。

そこで、より手軽な測定を実現するために、音声ファイルによる言語能力測定方法を提案する。手軽に音声録音可能なデバイスの一つに、スマートフォンがある。スマートフォンで録音した音声ファイルにより言語能力測定ができれば、より手軽に測定できると考えられる。本稿では、スマートフォンで録音した音声ファイルにより言語能力測定の可能性を検証する。

2 検証実験

スマートフォンで録音した音声ファイルによる言語能力測定の可能性を検証するため、実験を行った。本章では、実験内容と評価用データの作成について述べる。

2.1 実験内容

本実験では、Android 端末 (Galaxy J SC-02F) および iPod touch を用いて音声ファイルを録音する。以下、両端末をまとめて「スマートフォン」と表記する。実験協力者は日本語母語話者である大学生 10 名 (男性 5 名、女性 5 名、平均年齢 21.5 才) である。

録音する発話内容は、表 1 に示す 6 つのテーマから選んでもらうことにした。これらのテーマは、スピー

表 1: 発話のテーマ

テーマ
今後、どのような旅行をしてみたいか？
どういう家に住んでみたいか？
アルバイトをしていた時の思い出
旅行の思い出 (素晴らしいこと、失敗談など)
無人島に 3 つだけ持って行けるとしたら何を持っていくか？
1000 万円を 1 日で使うなら、どんなことに使うか？



図 1: 実験におけるスマートフォンの設置例

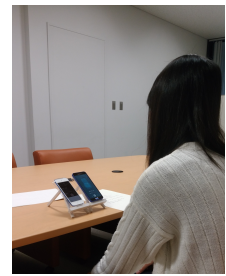


図 2: 実験の様子

チのネタ道場³のテーマから 3 分間程度発話できそうなものを事前調査し、上位 6 件となったテーマである。

実験では、各実験協力者に 6 つのテーマの中から 2 つテーマを選んでもらい、スマートフォンに向かって 1 つのテーマについて 3 分程度発話してもらった。

録音する前に、実験協力者にスマートフォンを持ってもらい、音声関連のソフトを利用する際のスマートフォンの位置を調べた。座った際にその位置となるようにスマートフォンを固定した上で、音声を録音した。本来は、手で持って録音することを想定しているが、今回の実験では Android 端末と iPod touch の 2 台で同時に録音するため、机の上に固定した。図 1 にスマートフォンの設置例 (左側が iPod touch, 右側が Android 端末) を、図 2 に実験の様子をそれぞれ示す。今回の実験は、録音を妨げる音が入らないようにするため、静かな環境で行った。

2.2 評価用データの作成

音声録音したデータ (Android 端末, iPod touch それぞれ 20 件) について、評価用データとして書き起こしデータ (以下、「書き起こし」と表記) および音声認識データ 2 種類の計 3 種類のテキストデータを用意した。書き起こしおよび認識結果 2 種類の例を表 2 に示す。

書き起こしデータは、録音データの発話内容を人手で正確に書き起こし、作成した。なお、「えーと」や「まあ」などの言いよどみ (フィラー) もテキストとして書き起こした。

音声認識データについては、音声認識システムによる認識精度の差異を想定し、2 つの音声認識システムを利用して作成した。今回は以下の 2 種類のシステムを用いた。

- 大語彙連続音声認識エンジン Julius [5] (以下, Julius の認識結果を「Julius」と表記)
- アドバンスド・メディア社の AmiVoice SP2⁴ (以

Verification of Measurement Possibility of Language Ability using Voice File

Hayato Choda[†] Eiji Aramaki^{††} Mai Miyabe[†][†]Faculty of Systems Engineering, Wakayama University^{††}Nara Institute of Science and Technology¹http://www.jasso.go.jp/about/statistics/intl_student_e/2015/index.html²http://www.kanken.or.jp/bjt/survey_reports/data/survey_reports.report01.pdf³<http://www.how2speech.com/minutes/>⁴<http://sp.advanced-media.co.jp/>

表 2: 録音データをもとに作成した評価用データの例

	書き起こし	Julius	AmiVoice
Android 端末	えっと今後どのような旅行をしたいかっていうと、とりあえず生きてる間にできるだけいろんな場所に行ってみたいです。	せっかくを、ノソノの旅行をしたいが、手入れ等、とりあえず生きてる間に、できるだけ、色んな場所にてです、	で、今後どのような旅行をしたいかっていうととりあえず言ってる間にできるだけいろんな場所に行ってみたいです
iPod touch		摂津今後が、どんなな旅行をしたいかっていうとお、お、お、おつゆ随時てりや入れに、できるだけ、いろんな場所にけりレス、	ヘッドホンのどのような旅行をしたいかっていうととりあえず生きてる間にできるだけいろんな場所に出れる

Julius, AmiVoice のテキストについては、音声認識システムから出力された結果をそのまま掲載している。

表 3: 各デバイスにおける TTR スコア

		平均	S.D.
書き起こし		0.430	0.059
Android 端末	Julius	0.614	0.046
	AmiVoice	0.564	0.038
iPod touch	Julius	0.587	0.043
	AmiVoice	0.588	0.041
All	Julius	0.600	0.047
	AmiVoice	0.576	0.041

All は Android 端末と iPod touch のデータの合計である。

表 4: 評価用データ間における TTR スコアの相関

	書き起こし /Julius	書き起こし /AmiVoice	Julius /AmiVoice
Android 端末	0.427	0.368	0.819
iPod touch	-0.038	0.454*	0.537*
All	0.193	0.396*	0.540*

* : $p < 0.05$

相関係数が 0.4 以上のものを太字で示している。All は Android 端末と iPod touch のデータの合計である。

表 5: 各発話者における TTR スコアの相関

実験協力者	書き起こし /Julius	書き起こし /AmiVoice
A	0.828	0.759
B	0.076	0.752
C	0.964	-0.612
D	-0.180	0.922
E	-0.869	-0.517*
F	-0.305	0.577
G	0.481*	0.966
H	0.257	0.855
I	-0.027	-0.017
J	-0.184	0.667

* : $p < 0.05$

相関係数が 0.4 以上のものを太字で示している。

下, AmiVoice SP2 の認識結果を「AmiVoice」と表記)

3 実験結果と考察

本研究では、認識誤りに頑健な指標である TTR に基づき、言語能力測定可能性を検証する。TTR は、Type (異なり語数) と Token (延べ語数) の比率 (Type/Token) であり、値が大きいほど語彙量が多いことを意味する。今回は、一つの発話内容 (3 分程度) ごとに TTR スコアを計算する。

表 3 に各デバイスの平均 TTR スコアを示す。表 3 より、デバイス間での大きな違いはみられなかった。つまり、測定するスマートフォンによる大きな影響はないと考えられる。

表 4 に評価用データ間の TTR スコアに関するピアソンの相関係数を示す。これらは、デバイス別にそれぞれ 20 件分 (10 名 × 2 発話)、All については 40 件分の発話データから相関を調べたものである。表 4 より、デバイス別にみると、Android 端末は書き起こしと Julius, iPod touch は書き起こしと AmiVoice の間に中程度の相関 (相関係数が 0.4 以上) がみられた。

表 5 に各発話者における TTR スコアの相関を示す。このデータは、実験協力者ごとに 4 件の発話 (Android 端末 2 件, iPod touch 2 件) のデータを用いて相関を調べたものである。表 5 より、特に AmiVoice に関しては、10 名中 7 名の協力者において、中程度～強い相関がみられた。

今回得られた結果を先行研究 [4] での結果と比較すると、評価用データ間における TTR スコアの相関 (表 4) および発話者における TTR スコアの相関 (表 5) のいずれも、同じ傾向が得られていることが確認できおり、スマートフォンで録音した音声データでも、言語能力を測定できる可能性があると考えられる。また、Android 端末と iPod touch に大きな差異はみられなかった。デバイスによる大きな影響は受けない可能性が高いが、デバイスごとに音声認識システムを変更することでより精度の高い認識ができる可能性がある。ただし、本稿で用いたデバイスは Galaxy J SC-02F および iPod touch のみであるため、今後他の機種についても問題なく測定できるかどうかを検証する必要がある。

4 おわりに

本稿では、スマートフォンを用いて録音した音声ファイルによる言語能力測定の可能性を検証した。2 種類のデバイス (Android 端末, iPod touch) および 2 種類の音声認識システムを用いた評価実験を行い、以下の 2 点を明らかにした。

1. 協力者ごとの Type・Token 割合を調べた結果、書き起こしと音声認識結果に正の相関が見られており、スマートフォンで録音した音声データでも言語能力を測定できる可能性がある。
2. デバイスによる言語能力測定への大きな影響はない可能性が高い。

本実験は、静かな環境で録音を行ったため、今後、別の環境で録音を行った場合でも同様の結果が得られるか検証する必要がある。

参考文献

- [1] 山口壘：日本企業における留学生人材の活用と労働市場での位置づけ，法政大学比較経済研究所，pp.1-21 (2016)。
- [2] Kemper, S., Marquis, J. and Thompson, M. : Longitudinal change in language production : effects of aging and dementia on grammatical complexity and propositional content, *Psychology and Aging*, Vol.16, No.4, pp.600-614 (2001)。
- [3] 北村育子：特別養護老人ホームに暮らす認知症利用者のアクセシメントの実際とその重症度評価における主要評価項目，日本福祉大学社会福祉論集，pp.33-46 (2006)。
- [4] 宮部真衣, 四方朱子, 久保圭 ほか：音声認識による認知症・発達障害スクリーニングは可能か？— 言語能力測定システム“言秤”の提案—, ワークショップ 2014 (GN Workshop 2014) 論文集, Vol. 2014, pp.1-8 (2014)。
- [5] 河原達也, 李晃伸：音声認識ソフトウェア Julius, *人工知能学会誌*, Vol.20, No.1, pp.41-49 (2005)。