

推薦論文

音声認識で生成した英語字幕による 英語理解向上の測定実験

下 郡 信 宏^{†1} 坪 井 創 吾^{†1,*1}

英語を母国語としない人が英語で会話を行う際に、音声認識によって自動生成された英語字幕を提示すると、理解度が向上するか実験により調べた。55名の英語能力の異なる被験者に対しTOEICのリスニング試験を、音声認識特有の誤りを含む字幕を提示しながら行った。その結果、TOEICレベルC(470~730点)の被験者に単語正解精度80%以上の字幕を提示した場合に成績が有意に向上した。TOEICレベルCは日本人受験者の約半数が属するレベルであり、単語正解精度80%は現在の音声認識システムを注意して使用すれば達成可能な精度であることから、音声認識によって自動生成された英語字幕でも会話の理解を支援することが十分に可能であることが確認できた。

Measuring the Effect of ASR Generated English Captions

NOBUHIRO SHIMOGORI^{†1} and SOUGO TSUBOI^{†1,*1}

We conducted an experiment to determine whether caption created by automatic speech recognition (ASR) will help non-native English speakers understand English conversation. 55 subjects with different English skill took the TOEIC listening tests while showing caption which includes recognition errors typical to ASR. There was significant difference when captions with accuracy more than 80% was presented to subjects with level C (score 470-730) on the TOEIC proficiency scale. Since about 50% of the Japanese TOEIC examinee fall under level C, and accuracy 80% can be achieved by current ASR technology, we conclude that caption generated by ASR is useful to non-native English speakers to understand English conversation.

1. はじめに

英語を母国語としない人が英語で会議を行うには多大な苦勞をとまなう。会議の内容を理解するのに余計に時間を要したり、会議の内容が正しく伝わらなかったりする場合もあり、母国語での会議に比べて非効率になりがちである。国際会議のような大きな会議であれば通訳を介して各出席者が母国語で円滑に会議を進めることもできるが、一般のビジネスでの会議ではコストがかかりすぎるため通訳を利用するのは難しい。グローバル化が進み、海外とのやりとりが増加するなか、英語を母国語としない人が英語で会議を行いやすくする安価な支援システムが求められている。

このような問題を解決するため、機械に通訳をさせる音声翻訳の研究も行われている^{1),2)}。音声翻訳では各人の発言を音声認識で機械に認識させた後に機械翻訳で対話相手の言語に翻訳する。音声翻訳では、音声認識の際に発生した誤認識を含む文を翻訳するため、機械翻訳による誤りも含めて、最終的な出力に多くの誤りが含まれやすくなる。このため、ビジネスでの実用に耐えるほどに十分な精度は得られていない³⁾。

そこで、我々は英語字幕に着目した。英語での発言を音声認識システムでテキスト化して英語のまま表示するシステムを検討している。英語での発言をテキスト化することで、英語で聞き逃した箇所を文字で確認できるため、理解度が向上することが期待できる。また、翻訳せずに英語のまま利用するため、機械翻訳の精度は問題にならない。さらに、英語を母国語とする人が自分の発言をテキストで確認して、不明瞭だった箇所を認識してはっきり言い直すことで、コミュニケーションが容易になることが期待できる。

しかし、音声認識システムの認識精度には限界があり、講義や会議における発言の認識率が近い将来に劇的に向上するのは当面は難しい。誤認識を含む字幕であっても、理解の支援として有効であるか確認する必要がある。

本研究では外国人、特に日本人であっても英語字幕があれば英語の理解度が向上することを確認する。さらに、音声認識特有の誤認識が含まれていても英語字幕があった方が英会話の理解度が向上するユーザ層が存在することを確認し、作成するシステムの目標性能を得る

^{†1} 株式会社東芝研究開発センター

Corporate Research & Development Center, Toshiba Corporation

*1 現在、東芝ソリューション株式会社プラットフォーム事業部

Presently with Platform Solutions Division, Toshiba Solutions Corporation

本論文の内容は2009年7月のマルチメディア、分散、協調とモバイル(DICOMO2009)シンポジウムにて報告され、同プログラム委員長により情報処理学会論文誌ジャーナルへの掲載が推薦された論文である。

ための実験を行った。

2. 関連研究

英語を母国語としない人に、音声と一緒に英語の字幕を表示すると理解度が向上することは以前より知られている。Garza は第二外国語として英語を学習している成人を観察し、字幕によりリスニングの理解度が向上すると報告している⁴⁾。また、Huang らも英語を第二外国語とする学生を対象にテレビの字幕 (Closed Caption) がリスニングの理解度に与える影響を調査し、リスニングの理解度が向上するだけでなく、語彙獲得能力も向上していると報告している⁵⁾。しかしこれらの研究では誤りが含まれない完全な字幕を用いているため、音声認識によって自動生成された、誤りを含む英語字幕でも同様にリスニング能力が向上するかは明らかではない。また、テレビ放送などの字幕は、一般的に発話が始まる直前、または発話開始と同時に表示される。このため、発話を聞きながら字幕を追いかけることが可能である。しかし、音声認識を用いて生成した字幕は発話が終わらないと表示されない。したがって、利用者には字幕が遅れて表示されるように見える。このような状況で理解度が向上するかも明らかではない。

音声認識を用いて聴覚障害者のために字幕を自動生成する研究が従来より行われている。たとえば、音声認識を用いてテレビ放送用の字幕を生成する研究が行われてる⁶⁾。松井らは発話内容を音声認識システムが認識しやすいように言い直すリスピーク方式により実用に耐える方式を提唱している⁷⁾。さらにニュース原稿が利用できるアナウンサの発言はそのまま音声認識を利用し、インタビューの部分など音声認識だけでは誤認識が増える部分はリスピーク方式で字幕を生成するハイブリッド方式なども提案されている。これらのシステムは放送に利用できる高品質な字幕を早く、簡単に生成することを目的にしている。しかし、訓練されたリスピーカなどを必要としており、認識精度は高いが、一般のビジネスシーンで利用できるほど手軽ではない。

もっと手軽なシステムとしては聴覚障害者が高度な専門教育を受けられるように、大学の講義の内容を音声認識で表示する試みが行われている。Liberrated Learning Project (LLP) では、複数の話者の発話を音声認識でテキスト化して表示するシステムを開発している⁸⁾。

類似するシステムとして、Munteanu らは録画された講義を後から検索して利用しやすくするために、音声認識を利用している^{9),10)}。この研究の中で、録画された講義の内容を音声認識で書き起こしたテキストを付加してビデオとともに学生に与えた。筆者らは、録画された講義時間よりも短い時間で講義内容を理解し、試験問題に答える実験を行った。この中

で、Word Error Rate (WER) 25%以内なら音声認識によるテキストがあった方が成績が向上することが確認されている。

これらのシステムは、英語を母国語とした学生を対象としており、英語を母国語としない人が利用した場合の効果は明らかではない。したがって、以下の実験では、音声認識による誤りを含む英語字幕を、英語を母国語としない日本人に提示した場合に、理解度が向上するかを検証する。

3. 実験

字幕の有無や字幕の認識精度の違いによって英語での発言の理解度がどの程度変わるかを測定する実験を行った。実験では被験者に4つの字幕環境でTOEICのリスニング試験を受けてもらい、成績の変化を測定した。

TOEICのリスニング試験は実際の会議でのやりとりとは設定が異なるが、英語の理解度を測定するための試験として確立しているため、定量的に理解度を測定するには有効である。

実際の会議では聞き取り以外の要因が理解に影響を及ぼすが、TOEICを使用することでこれら、聞き取り以外の要因を排除することができるため、音声認識による英語字幕の効果測定する基礎的実験として有効な設定と考えた。

この実験で検証したい仮説は以下の3つである。

(仮説1) 英語があまり得意でない人に字幕を提示すると理解度が上がる。

(仮説2) 英語力の高い人は字幕を表示しても成績は変化しない。

(仮説3) 認識精度の悪い字幕を提示すると、逆に成績は悪くなる。

仮説1では英語字幕がプラスに働く人が存在すること、仮説2では英語力の高い人には英語字幕がマイナスに働かないこと、そして仮説3では認識精度が悪すぎるとマイナスに働く場合があることを確認している。すなわち、3つの仮説が支持されれば、一定以上の認識精度が達成できれば、英語字幕は役に立つことが検証できる。以上の3つの仮説を検証するために以下の実験を行った。

3.1 被験者

被験者は55名。TOEICを2年以内に受験し、トータルスコアが220以上の人を対象とした。スコア220以下はTOEIC公式のTOEIC PROFICIENCY SCALE¹¹⁾でレベルEに該当し、「コミュニケーションができるまでに至っていない」と判定される人であるため、被験者の対象から除外した。また、実験で使用する市販の問題集を使用していないことを確

認してある。被験者は TOEIC のレベルが A ~ D の 4 つのレベルで、各レベルの被験者数ができるだけ同数となるように集めた。被験者の TOEIC レベルは自己申告された点数に基づいて分類した。被験者には実験への参加報酬を支払った。被験者の属性は以下のとおりである。

レベル：A 11 名，B 13 名，C 16 名，D 15 名，E 0 名

性別：男性 33 名，女性 22 名

年齢：20 代 12 名，30 代 16 名，40 代 25 名，50 代 2 名

3.2 試験問題

試験問題には TOEIC の実施団体である Educational Testing Service が出版する最新の公式問題集¹²⁾を利用した。この問題集には TOEIC の模擬試験問題が 2 回分含まれている。2 回分の模擬試験のうち、Part 2 のリスニングの試験問題を 30 問ずつ、合計 60 問を使用した。TOEIC の Part 2 は短い文章が読み上げられ、その文への応答が 3 文読み上げられ、3 文の中から最初の文章への応答として最も相応しい文章を選択する形式の問題である。音声以外に手がかりはなく、会話を理解する能力を問われる試験であるため、会議での会話理解力を調べるには最も適している。

以下は試験問題の例である。以下の 4 文が音声のみで読み上げられる。

Did you make a dinner reservation?

(A) I prefer fish.

(B) Flight 261 to Osaka.

(C) Yes, it's at 7 o'clock.

上記の問題の場合、(C) が正解である。

3.3 字幕

60 問の試験問題を 15 問ずつ 4 つのグループに分割してそれぞれに認識精度の異なる字幕を付けて実験を行う。字幕は以下の 3 種類を用意した。

(1) 誤りを含まない精度 100% の字幕

(2) 誤りを 20% 含む精度 80% の字幕

(3) 誤りを 30% 含む精度 70% の字幕

さらに、基準とするために通常の TOEIC 試験と同じ環境である「字幕なし」を含めて 4 種類の字幕環境を使用した。

試験問題の 4 つのグループに 4 種類の字幕環境を用意し、試験問題のグループと字幕の種類の組合せが被験者ごとにできるだけ重複しないようにした。字幕の精度は音声認識の研究

で一般的に用いられている以下の正解精度を利用した：

$$\text{正解精度} = (\text{正解単語数} - \text{挿入誤り数}) \div \text{全単語数} \times 100 \quad (\%)$$

字幕に含まれる誤りは音声認識システムが起こす誤認識と同様の誤りを混ぜるために、音声認識システムに問題文と選択肢を読み上げて音声認識させることで再現した。読み上げは英語を母国語とする人を用意できなかったため、TOEIC 950 点のネイティブに近い人で代用した。音声認識システムにはドラゴンスピーチ 2005 (Dragon Naturally Speaking 2005)¹³⁾を使用した。付属のトレーニングツールで、事前に音声認識システムの辞書を、2 時間程度学習させている。音声認識により得られた誤りの混じったテキストは精度が 70% 以下であったため、誤認識した箇所を正解に戻すことで、正解の量を調整して 15 問の合計で精度 70% と精度 80% の字幕を作成した。なお、調整に際して、間違いに偏りが起きないように、1 問の中に間違いが最低 1 つ存在するようにした。1 問とは問題文と回答の選択肢の 3 文の合計 4 文からなる。

以下が問題のサンプルである。

精度 100%

Don't you need a ticket for the show?

(A) I already have one.

(B) Yes, I think it might snow.

(C) I took the train.

精度 80%

Don't you need a ticket for the show?

(A) I already have on.

(B) Yes, I think it might snow.

(C) I took the train.

精度 70%

Don't you need a ticket for the show

(A) I already have on.

(B) Yes, I think it might snow.

(C) I looked the train.

注) 下線は誤認識箇所を表すが、実験時に被験者には提示していない。

被験者はヘッドホンを使用して問題文を聞いている。被験者の前にはディスプレイが設置されており、字幕を提示する設定では問題文が 1 文、ヘッドホンから流れ終わると、時間を



図 1 実験風景
Fig. 1 Experiment environment.

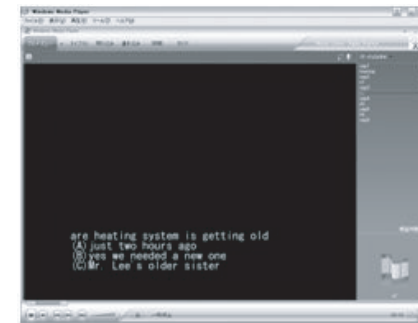


図 2 字幕表示画面
Fig. 2 Caption display screen.

おらずに字幕が表示される。ただし、通常の映画などの字幕とは違い、1文を読み終わるまでは字幕は表示されない。これは音声認識システムを利用した字幕システムの実際の動きを再現するためである。

字幕は1行40文字、最大4行まで表示される。新たに行を追加して表示すると、上にスクロールして一番上の行が消える。

3.4 実験手順

被験者は最初に15問の練習問題を行う。練習では字幕なし3問、字幕精度100%4問、字幕精度80%4問、字幕精度70%4問の順に問題が提示される。字幕の精度が切り替わるたびに字幕生成システムが変わったと利用者には画面上で通知した。これにより字幕生成システムにより誤認識の量が異なることを意識させた。

練習が終わると、音量や座席の最終調整と質問のための時間がとられる。続いて、本番の実験に入る。15問の問題を字幕なし、字幕精度100%、字幕精度80%、字幕精度70%のいずれかの字幕環境で行う。これを1セットと呼ぶ。1セット終了後にアンケートに回答する。アンケート終了後に次のセットを異なる字幕環境で行う。以上を繰り返し、4セットで4種すべての字幕設定を完了させる。後の集計で使用するために、各問題を解答すると同時に、字幕を参考にした場合には解答用紙の各問の回答欄の横に設けられたチェックボックスに印を入れるように指示した。実験風景を図1に、字幕表示画面を図2に示す。

3.5 アンケート

各セット終了後に被験者の主観的な評価を取得する目的で以下のようなアンケートを行った。

- 字幕を参考にしましたか

- (1) 毎回参考にした
- (2) ときどき参考にした
- (3) あまり参考にしなかった
- (4) まったく参考にしなかった

- 字幕は正確でしたか

- (1) とても正確だった
- (2) ほぼ正確だった
- (3) あまり正確でなかった
- (4) まったく正確でなかった
- (5) 分からない

また、4セットすべてが終了した後に以下のアンケートを行った。

- 字幕は役に立ちましたか

- (1) とても役に立った
- (2) ある程度役に立った
- (3) あまり役に立たなかった
- (4) まったく役に立たなかった(必要なかった)

- 英語の会議に出席した経験はありますか

- (1) ない
- (2) ある

- 英語の会議でこのように字幕が出るとしたら役に立つと思いますか
 - (1) とても役に立つと思う
 - (2) ある程度役に立つと思う
 - (3) あまり役に立たないと思う
 - (4) まったく役に立たないと思う

4. 結 果

4.1 仮説の検証

3つの仮説の結果を確認する。

(仮説1) 英語があまり得意でない人に字幕を提示すると理解度が上がる。

図3はTOEICレベル別、字幕環境別の成績の平均を表すグラフである。各セット15問であるため15点満点である。字幕をまったく参考にしなかったとアンケートで回答した被験者を除いてある。このため集計数は被験者数と一致しない。字幕を参考にしなかった場合は、字幕の影響が表れずノイズとなってしまうため、字幕を1度も参照しなかったと回答した被験者は集計から取り除いた。

図より、レベルA、Bの被験者グループでは字幕の有無にかかわらず正解数が13~14と高いことが分かる。これらの被験者グループでは字幕による有意な差は見られない。一方、レベルCの被験者グループでは、字幕なしと字幕精度80%で有意水準1%で有意差があり、

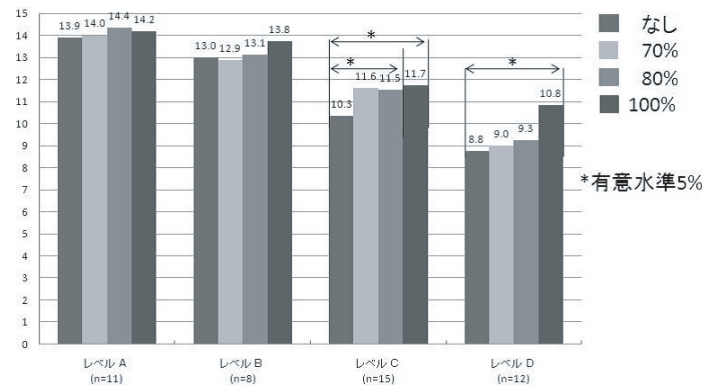


図3 レベル別、字幕環境別の成績

Fig. 3 Test score by skill level and recognition accuracy.

また、字幕なしと字幕精度100%で有意水準5%で有意差が確認できた。したがって、あまり英語が得意でないTOEICレベルCの人に精度80%以上の字幕を提示すると理解度が上がることが確認できる。これにより仮説1は支持された。

加えてTOEICレベルDの被験者グループでは字幕なしと字幕精度100%で有意水準5%で有意差が確認できたが、字幕なしと字幕精度80%では有意差がなかったため、レベルDの人は認識精度80%よりも高い精度の字幕でなければ理解度が向上しないことが分かる。

続いて、仮説2を検証する。

(仮説2) 英語力の高い人は字幕を表示しても成績は変化しない。

すでに説明したとおり、図3において、レベルAとレベルBの被験者グループはいずれの字幕精度でも有意差が見られない。したがって、レベルB以上に英語力の高い人は字幕を表示しても成績は変化するとは言いえない。よって仮説2も支持された。これにより精度が悪い字幕を提示してもレベルB以上に英語力の高い人の理解力に悪影響を与えることがないことが分かる。

最後に仮説3を検証する。

(仮説3) 認識精度の悪い字幕を提示すると、逆に成績は悪くなる。

いずれのレベルの被験者グループでも字幕なしと字幕精度70%では有意差は見られない。したがって、70%の認識精度では悪影響があるとは言いえない。今回の実験では仮説3は支持されなかった。

4.2 アンケート結果

各セット終了後に行ったアンケートの結果についてまとめる。回答用紙の回答欄の横に、問題ごとに字幕を参考にしたか否かを記すチェックボックスが設けてあり、字幕をまったく参考にしなかった被験者を集計結果から除外するのに用いた。また、被験者が英語を理解するのに字幕を参考にしていることを確認するために字幕を参考にしたか否かを聞いている。

図4はレベル別に字幕を参考にした数の平均を表している。この結果からTOEICのレベルが低いほど、字幕を参考にする割合が多いことが分かる。また、字幕精度は100%だと参考にされる割合が高まる。図5はアンケートから得た字幕を参考にした度合いの結果である。図4に示す実際の参照数と同様にTOEICのレベルが低いほど、また字幕精度が高いほど参考にする割合が高まっており、これは仮説1を裏付ける結果となっている。

次にすべてのセッションが終わった後に行った最終アンケートの結果についてまとめる。被験者の英語字幕の有用性に関する主観評価が試験の成績と乖離していないことを確認するために、アンケートを行っている(全体を通して字幕は役に立ちましたか)。

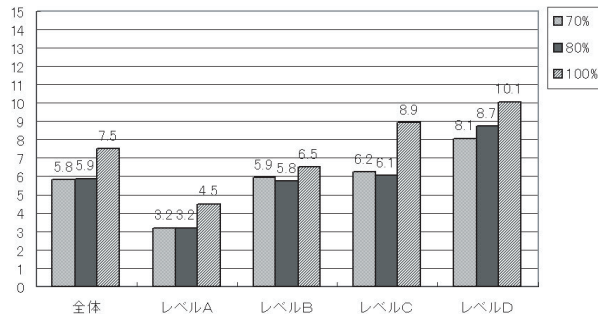


図 4 レベル別字幕参照数の平均

Fig. 4 Actual times caption was used.

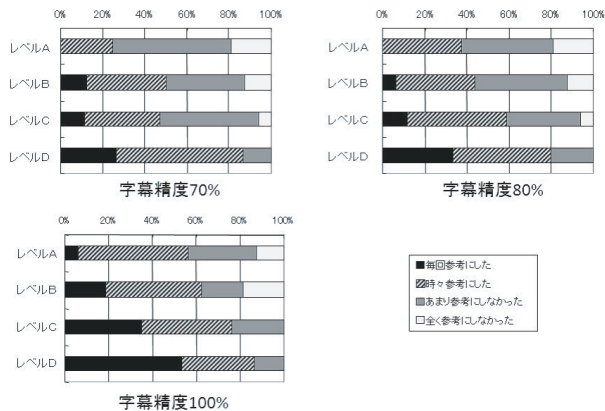


図 5 字幕参照度合い

Fig. 5 Caption usage.

「とても役に立った」「ある程度役に立った」を合わせると全体では 6 割弱が役に立ったと回答した。傾向として TOEIC のレベルが低いほど「役に立った」と回答する割合が高い。レベル A では「とても役に立った」との回答が 0 件であるのに対し、対称的にレベル D では「まったく役に立たなかった」が 0 件であった。この結果も仮説 1 と仮説 2 を支持するものとなっている (図 6)。

役に立った/立たなかった理由を見ると、役に立たなかった理由としては、字幕システム

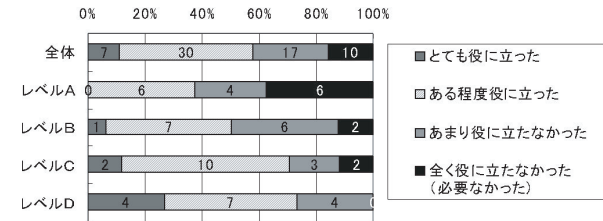


図 6 全体を通して字幕は役に立ちましたか

Fig. 6 Was the caption useful?

に起因するもの (タイミング, 精度など), 英語レベルに起因するもの (問題が簡単で不要, 見ても意味をつかめないなど), および, 字幕の利用そのものについて (見る/聞くが両立できない, 読むのは時間がかかるなど) の内容がある。

- 役に立った
 - 聞き逃した単語を確認できる
 - 内容を理解するのに参考になった
- 役に立たなかった
 - 間違いがある/間違いがあると混乱する
 - 気が散る/聞く方に集中できない
 - 聞く・見るが両立できない
 - 読むのは時間がかかる
 - タイミングが遅い
 - 問題が簡単なので不要 (レベル A)
 - 難しい英文だと見てもすぐに意味をつかめない (レベル D)

(字幕は正確でしたか)

字幕がどの程度正確と感じたかを比較すると図 7 のようになる。字幕精度 80% で TOEIC レベル C, D の人の中に「とても正確だった」と回答している人が存在する。字幕が「とても正確であった」, または「ほぼ正確であった」とするレベル D の被験者の割合は字幕精度 70% と 80% であまり差がないが, レベル C の被験者では「とても正確であった」とした被験者が 0 人であり正確であったとする被験者の割合が大きく減る。レベル C の被験者は字幕の間違いに気づき, 間違っていない箇所だけを選び分けて利用できていると考えることができる。

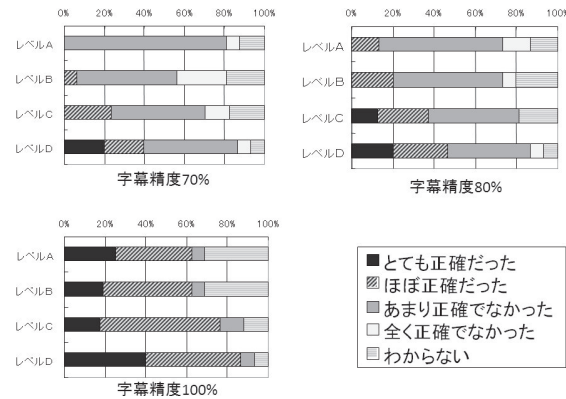


図 7 字幕は正確でしたか

Fig. 7 Was the caption accurate?

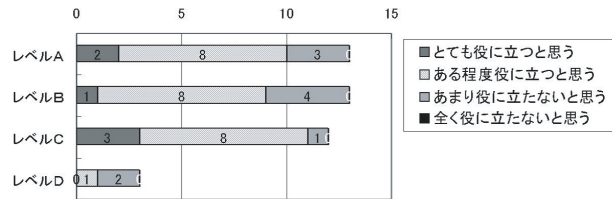


図 8 英語の会議で役に立つと思いますか

Fig. 8 Will captioning help in an English meeting?

英語会議経験者に「英語の会議でこのように字幕がでると役に立つと思うか」を聞いた。全体では、「とても役に立つと思う」「ある程度役に立つと思う」合わせて 8 割弱が役に立つと回答した。逆に「まったく役に立たないと思う」の回答は 0 件であった (図 8)。

個別の意見を見ると、役に立つ理由としては、TOEIC テスト時と同様、

- 聞き逃したところの確認に役立つ
 - 出席者に英語力のばらつきがあるので補助機能として
- などの意見があったが、「正確であれば」の条件付きのコメントも多い。役に立たない理由、懸念点、要望としては、
- 間違いが多いと誤解につながる
 - タイミングが悪い

- 非ネイティブの国では精度が悪くなるのでは
- その場で聞き返せる状況であれば不要
- 表示行数を多くしてほしい
- 文書化、議事録として利用できるとよい

などがあげられており、認識精度に対する懸念が強いことが分かる。

5. 考 察

実験の結果より TOEIC レベル C の被験者に正解精度 80% の字幕を提示すると成績が向上することが確認できた。TOEIC のレベル C は日本の受験者の約 50% が該当するレベルである¹⁴⁾。また、字幕の正解精度 80% とは現状の音声認識システムで英語を母国語とする人がヘッドセットなどを利用した理想的な音環境で丁寧に話せば達成可能な認識精度である。TOEIC の受験者の母体が英語で会議をする人の母体と一致すると仮定すれば、現行の技術で英語字幕システムを提供すれば日本人の約半数に効果があるシステムが提供できることが分かった。

今回の実験ではレベル D は字幕精度 100% のみ効果が現れたが、字幕精度 80% と 100% の間で効果がある可能性も残っている。今回の実験で、字幕精度 90% の問題を作ろうとすると、誤認識を含まない問題ができてしまうため、字幕精度 90% の実験を行うことができなかった。異なる問題を利用して、エラーを一様に分散させつつも字幕精度 90% の問題を作成できるように実験方法を工夫する必要がある。

被験者のレベル分けは自己申告に基づいて行ったが、これがノイズとなった可能性がある。また、TOEIC の総合点ではなく、リスニングの点数だけであらかじめレベル分けを行った方が、精度が高まった可能性がある。

認識精度 70% (WER 30%) では効果がなかったが、認識精度 80% (WER 20%) では効果があったことは、Munteanu らの結果⁹⁾ の WER 25% 以内ならば有効とした結論と一致する。英語を母国語としない被験者であり、また、タスクも異なるにもかかわらず同じ結果が出たことに関して、今後さらに調べる必要がある。

レベル C で字幕精度 100% の点数とレベル B で字幕なしの点数ではレベル B の平均点の方が高いことが興味深い。残念ながら、有意差は得られていないが、ここに差が確認できれば、レベル C の利用者には完璧な英語字幕を提供してもレベル B の能力を得ることはできないことを意味しており、レベル B の能力にまで押し上げるには、翻訳機能を提供しなければならないことになる。自己申告によるレベル分けがノイズとなっている可能性があるた

め、今後さらに詳しく実験していきたい。

最後に、図7において、字幕精度が100%であるにもかかわらず、「とても正確だった」とした人がレベルDを除いて30%に満たない。被験者が単に最高点を付けるのを躊躇しただけか、字幕の表示タイミングが遅いために減点されたものか、もしくは字幕の精度をいくら向上させても利用者を完全に満足させることができないことを意味しているのか、さらなる検証が必要である。

6. おわりに

80%の認識精度の字幕を提示することでTOEICレベルCの被験者の理解度が向上することが確認できた。認識精度80%とは現行の音声認識システムでも英語を母国語とする発話者がヘッドセットなどを使った良好な音質で、さらにシステムが理解しやすいように丁寧に発話すれば実現可能な精度である。また、TOEICレベルCは日本のTOEIC受験者の約50%が属するレベルである。したがって英語を母国語とする人との英語での会議において、音声認識による英語字幕による支援が、有効であるターゲットユーザが存在することが確認できた。実際の会議は今回の実験と異なり双方向のやりとりが発生するため結果も異なると考えられる。今後は、音声認識を用いた字幕システムを実際の会議で試用し、その効果を検証していく予定である。

参考文献

- 1) 竹沢寿幸, 森元 暉, 匂坂芳典, キャンベルニック, 飯田 仁: 日英音声翻訳システムATR MATRIX, 情報処理学会全国大会講演論文集, Vol.56, No.2, pp.279-280 (1998-03-17).
- 2) 知野哲朗, 釜谷聡史, 降幡建太郎, 住田一男: 日中英3言語6方向音声翻訳システム(翻訳・要約), 情報処理学会研究報告, SLP, 音声言語情報処理, Vol.2008, No.46, pp.15-22 (2008-05-15).
- 3) 菅谷史昭, 竹沢寿幸, 横尾昭男, 山本誠一: 音声翻訳システムと人間との比較による音声翻訳能力評価手法の提案と比較実験, 電子情報通信学会論文誌 D-II, 情報・システム, II-パターン処理, Vol.84, No.11, pp.2362-2370 (2001-11-01).
- 4) Garza, T.J.: Evaluating the Use of Captioned Video Materials in Advanced Foreign Language Learning, *Foreign Language Annals*, Vol.24, pp.239-258 (2008).
- 5) Huang, H. and Eskey, D.E.: The Effects of Closed-Captioned Television on the Listening Comprehension of Intermediate English as a Second Language (ESL) Students, *Journal of Educational Technology Systems*, Vol.28, No.1, pp.75-96 (1999).
- 6) 安藤彰男, 今井 亨, 小林彰夫ほか: 音声認識を利用した放送用ニュース字幕制作シス

テム, 電子情報通信学会論文誌 D-II, 情報・システム, II-パターン処理, Vol.84, No.6, pp.877-887 (2001-06-01).

- 7) 松井 淳, 本間真一, 小早川健ほか: 言い換えを利用したスピーク方式によるスポーツ中継のリアルタイム字幕制作(音声, 聴覚), 電子情報通信学会論文誌 D-II, 情報・システム, II-パターン処理, Vol.87, No.2, pp.427-435 (2004-02-01).
- 8) Wald, M. and Bain, K.: Using Speech Recognition for Real-Time Captioning of Multiple Speakers, *IEEE MultiMedia*, Vol.15, pp.56-57 (2008).
- 9) Munteanu, C., Baecker, R., Penn, G., Toms, E. and James, D.: The effect of speech recognition accuracy rates on the usefulness and usability of webcast archives, *CHI '06: Proc. SIGCHI Conference on Human Factors in Computing Systems*, pp.493-502, ACM, New York, NY, USA (2006).
- 10) Munteanu, C., Baecker, R. and Penn, G.: Collaborative editing for improved usefulness and usability of transcript-enhanced webcasts, *CHI '08: Proc. 26th Annual SIGCHI Conference on Human Factors in Computing Systems*, pp.373-382, ACM, New York, NY, USA (2008).
- 11) 国際ビジネスコミュニケーション協会: TOEIC PROFICIENCY SCALE (2010). <http://www.toeic.or.jp/toeic/pdf/data/proficiency.pdf>
- 12) Educational Testing Service: TOEIC テスト新公式問題集 Vol.3, 国際ビジネスコミュニケーション協会 (2008).
- 13) Ascendia Inc.: ドラゴンスピーチ 2005 (2005). <http://www.dragonspeech.jp/>
- 14) TOEIC 運営委員会: TOEIC テスト DATA & ANALYSIS 2008 (2008). <http://www.toeic.or.jp/toeic/pdf/data/DAA2008.pdf>

(平成22年1月10日受付)

(平成22年6月3日採録)

推薦文

日本人が外国人と英語で会議を行う場面で、音声認識による英語字幕を提示する効果を測定した研究である。音声認識によって自動生成した英語字幕は一般のテレビや映画などに表示される字幕と異なり、認識誤りや表示の遅延などが生ずるといった欠点がある。しかし、そのような欠点があっても、音声認識により生成した字幕を提示した方が会話の理解が有意に向上することを実験により明らかにした有用性に富んだ論文である。

(マルチメディア, 分散, 協調とモバイル (DICOMO2009) シンポジウム
プログラム委員長 土井美和子)



下郡 信宏（正会員）

1990年東京工業大学理学部情報科学科卒業。1992年東京工業大学大学院情報科学専攻修士課程修了。同年（株）東芝入社。1998～2000年スタンフォード大学コンピュータサイエンス学科客員研究員。ヒューマンインタフェース，ネットコミュニティシステム，コミュニケーション支援技術の研究に従事。



坪井 創吾（正会員）

1998年東京工業大学生命理工学部生物工学科卒業。2000年東京工業大学大学院総合理工学研究科知能システム科学専攻修士課程修了。同年（株）東芝入社。CSCW，コミュニティ支援，同期・非同期コミュニケーション支援システム等の研究に従事。人工知能学会会員。