6 英語字幕による会議支援

下郡信宏　（(株)東芝　研究開発センター）

普通の日本人にとって外国人と英語で会議を行うのは楽ではない．にもかかわらず，グローバル化により海外との交流は増え，通信技術の発達によりテレビ会議が手軽に利用できるようになったため，英語で会議をする機会は増える傾向にある．

そこで，すでにある技術を使って日本人の英語でのコミュニケーションを取る支援をする方法として音声認識を使って英語の字幕を表示することに着目し，その有効性を確認した実験を紹介する．

理想は吹き替え

おそらく，最も理想的な支援方法はテレビの吹き替えのように，相手がしゃべった英語を通訳して日本語で発話しなおしてくれる音声翻訳であろう．実際に，そのような研究はすでに行われている[1),2)]．しかし，音声翻訳は通常，英語の発言を音声認識によりテキスト化した後に，機械翻訳で英語を日本語に変換し，最後に日本語を音声合成で読み上げる，という 3 つの処理を連続して行っている．そのため，それぞれの処理で誤りが入る余地があるため，会議などのビジネスシーンで使える精度を得るのはハードルが高い．

たとえば，以下の発言があったとする．

"I think the accuracy is pretty good."

この文を機械翻訳で英日翻訳すると，「私は，精度がかなり良いと思います．」と正しく翻訳される．

しかし，最後の *"good"* が音声認識で正しく認識されず，*"goo"* と認識されてしまったとする．

"I think the accuracy is pretty goo."

英文の状態なら元の発言を推測することは可能だが，これを翻訳すると，「私は，精度がきれいなべとべとしたものであると思います．」と間違って翻訳されてしまう．音声認識で 1 文字聞き逃しただけなのに，もはや元の英文が何であったのか，想像すらできない．これではコミュニケーションを支援するどころか，逆に妨げになってしまう．現在の音声認識の精度は使用環境により大きく異なるが，会議などの場合は 80% 前後であるため，このような間違いは頻繁に発生する．したがって，音声翻訳を使ってビジネスができるようになるにはまだしばらく時間がかかりそうである．

確かに日本語に吹き替えられた状態での会話がベストかもしれないが，現状の技術でも何か支援ができるのではないかと考え，注目したのが英語字幕である．

英語での発言を音声認識でテキスト化して日本語に翻訳せずに英語のまま字幕として表示すると，聞き取りが難しかった個所を視覚的に確認することができるため，コミュニケーションが楽になるのではないかと考えた．

過去の研究により，英語の音声に英語の字幕を表示すると，理解しやすいことが実験で確認されている[3)]．この実験では英語を母国語としない生徒に英語のビデオ教材を見せた後に内容を確認するテストを行っている．その結果，英語字幕をつけて学習した生徒の方が英語字幕をつけずに学習した生徒よりもテストの成績が良かったため，英語字幕は効果があることが確認されている．

この考えを応用した英語学習方法が，英語のドラマを英語字幕を表示させながら視聴する方法である．英語字幕を表示しながら英語の音声を聞くと，聞き取りが容易になるため学習効果が高まると考えられる．このように英語字幕は英語の聞き取りを支援する手段としてすでに使われている．

まったく勉強したこともない外国語であれば，翻訳しないと理解できないが，英語に限っては日本語に翻訳せずに英語のままの字幕でも聞き取りを支援することが可能なのである．

テレビの字幕との違い

映画やテレビドラマの字幕と音声認識で生成した会議の字幕では 2 つの点で大きく異なる．1 点目は精度である．映画やテレビドラマの字幕に間違いはほとんど含まれないのに対して，音声認識で生成した字幕には誤認識を含む．

このため，誤った字幕を信じて相手の発言を誤解してしまったり，字幕の間違いに気を取られて会議に集中で

きなくなる可能性がある．

先ほどの例では，*"I think the accuracy is pretty goo."* と表示されたときに，*"goo"* とは *"good"* の間違いだな，と考えてしまい，余計な負荷がかかる．

2 点目はタイミングである．映画やテレビドラマの字幕は役者が発言するのと同時，またはワンテンポ早く表示されている．このため，耳で発言を聞きながら，同時に目で文字を追うことができる．一方，音声認識で生成した会議の字幕は相手が一文を発言し終わらないと認識が完了しない．このため，発言に遅れて字幕が表示されることになる．

したがって，音声をいったん理解した後に，再度字幕を確認するため，二度手間となり，利用者の負荷は高くなると考えられる．

このように，音声認識で生成した字幕は映画やテレビドラマの字幕と異なる性質を持っており，同じように役に立つと結論することはできない．

そこで，音声認識で生成した，誤りを含む字幕を，発言からワンテンポ遅れて表示しても理解の支援になるか，実験によって検証する必要がある．

英語字幕は役に立つか

◉ 実験の目的

今回の実験は以下の 3 点を確認することを目的としている：

1. どのレベルの英語力の人に有効か？

ネイティブ並みに英語が得意な人は字幕がなくても理解できるはずである．

英語があまり得意でない層に有効であると思われる．

2. 認識精度はどの程度必要か？

字幕の精度が悪すぎると，役には立たないのは明らかである．

どの程度まで，誤りを含んでいても役に立つのかを調べたい．

3. 認識精度が悪いと悪影響があるか？

字幕の精度が悪すぎると，役に立たないだけでなく，逆に理解を妨げていないかを確認したい．

◉ 実験の内容

実験は被験者 64 名で行った．被験者の英語力が TOEIC の公式のレベル分けでレベル A ～ D の 4 つのレベルでほぼ同数となるように揃えた [4]．

被験者は英語字幕を見ながら TOEIC のリスニングの試験を受け，字幕の違いによる成績の違いを比較した．試験問題は TOEIC の公式の問題集から Part 2 のみを使用した [5]．TOEIC の Part 2 は最初に短い文が読み上げられ，次にその文への応答が 3 文読み上げられ，3 文の中から最初の文への応答として最も相応しいもの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.*

（Educational Testing Service：TOEICテスト新公式問題集<Vol. 3>より引用）

上記の問題の場合，(C) が正解である．

字幕は正解精度 100%，80%，70% の 3 種類を用意し，さらに比較のため字幕がなく，音声だけの試験環境，すなわち通常の TOEIC 試験と同じ環境でも試験する．

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ている：

正解精度 =

（正解単語数－ 挿入誤り数）÷ 全単語数× 100（%）

音声認識による誤認識は人間の聞き間違いとは異なる間違え方をする．字幕に含まれる間違いに音声認識特有の誤認識を再現するため，市販の音声認識ソフトウェアに問題文を読み上げて認識させた上で，間違いの数を調整することで目的の精度の字幕を作成した．

以下が提示した字幕の例である．

[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.*

（Educational Testing Service：TOEICテスト新公式問題集 <Vol. 3> より引用）

[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）下線は誤認識個所を表すが，実験時に被験者に下線は提示していない．

被験者は字幕なしを含めた 4 種類の字幕環境で 15 問ずつ試験問題を解答する．

字幕の種類が切り替わるときには「字幕プログラム」が切り替わったと被験者に知らせることで，精度が悪い字

図 -1　実験の様子

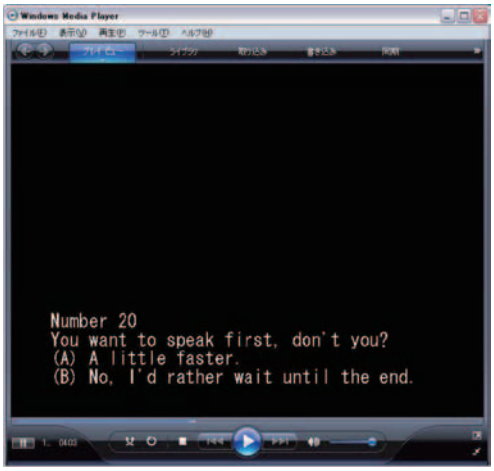

図 -2　字幕の表示例

幕を見た被験者が，それ以後のセットで字幕を参考にしなくなるのを防いだ．

図 -1 は実験の様子の写真である．被験者により試験問題の順番が異なるため，ヘッドホンを使用して，同時に受験している．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たびに，時間を置かずに字幕が表示される．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．

図 -2 は字幕の表示例である．画面には字幕以外は何も表示していない．

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．練習段階でも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．

◉実験の結果

図 -3 は被験者のレベル別に得点を集計したグラフである．レベル A は英語が得意な被験者グループでレベル D に向かって不得意になってゆく．

レベルごとに 4 つのバーが伸びており，左から字幕なし，精度 70%，精度 80%，精度 100% の得点を表す．

実験の結果，レベル C の被験者で字幕なしと精度 80% の字幕を提示した場合には有意水準 5% で有意差が確認できた．また，字幕なしと精度 100% の字幕を提示した場合には有意水準 1% でレベル C とレベル D の被験者で有意差が見られ，それ以外の組合せでは有意な差は見られなかった．

実験の目的としてはじめに挙げた疑問を順に確認してゆく．

・どのレベルの英語力の人に有効か？

上記の結果から TOEIC のレベル C の人に有効であることが分かる．

TOEIC のレベル C とは 470 ～ 730 点の人であり，日本での TOEIC 受験者の約半数が属するレベルである．これにより英語字幕で支援できるユーザ層が存在することが確認できた．

一方，レベル A やレベル B の英語力の高い被験者は精度 100% の字幕を提示しても成績は向上していないため，英語が得意な人には英語字幕は効果がないことが分かる．

・認識精度はどの程度必要か？

レベル C の被験者は精度 80% で成績が向上しているが，英語力の低いレベル D の被験者は精度 100% になってはじめて有意に成績が向上している．音声認識で精度 100% を達成するのは近い将来では難しいが，精度 80% であれば，現在の音声認識システムでも，はっきり喋るなど，注意して使用すれば達成可能なレベルであるため，レベル C のユーザを対象に精度 80% を目標とすることが適切と考えられる．

今回の実験では明らかになっていないが，精度 80% と精度 100% の間にレベル D のユーザにとって有効な精度が存在するかもしれない．これを明らかにするためにはさらに詳しい実験が必要である．

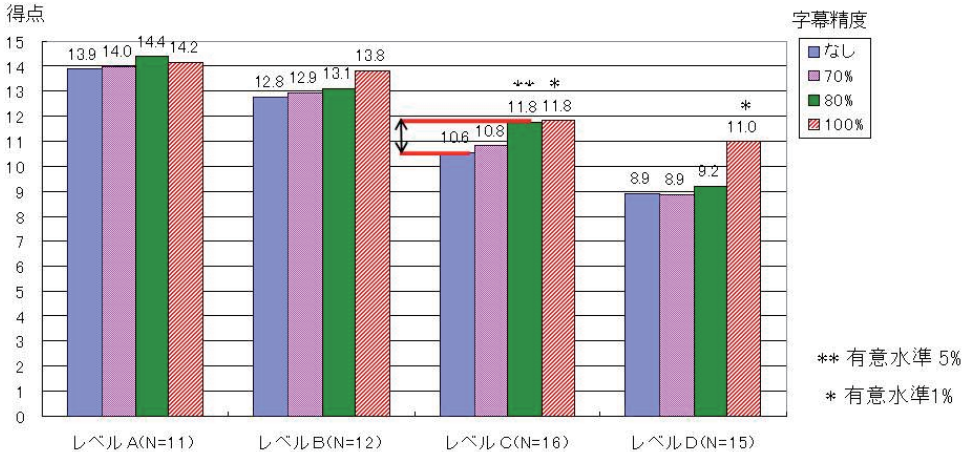

図 -3　レベル別の得点

・認識精度が悪いと悪影響があるか？

いずれのレベルの被験者でも字幕なしと精度 70% の字幕の間で有意な差はないため，精度 70% 程度ではいずれのレベルのユーザにも悪影響はないと言える．

ただし，精度 70% 以下の字幕を提示しても悪影響がないとは，今回の実験からは断言できない．もっと精度が下がれば悪影響が出る可能性を否定することができない．

役に立つシステムに向けて

今回の実験から，音声認識で生成した英語字幕でも精度 80% 以上であれば TOEIC レベル C のユーザには役に立つことが分かった．精度 80% は現在の技術でも達成可能であり，また TOEIC レベル C は日本の TOEIC 受験者の約半数が該当するレベルであることから，会議で使える音声認識による英語字幕システムの可能性を示すことができた．

ところが，英語字幕を表示した方が成績が向上しているにもかかわらず被験者は必ずしも満足していないという結果も得られている．図 -4 は被験者に各セット終了ごとに行ったアンケートの結果である．字幕がどの程度正確であったと思うか質問しているもので，被験者の字幕に対する満足度を表していると考えられる．このグラフは精度 100%，つまり誤りを含まない完璧な字幕を表示した場合の結果をまとめたものであるが，精度 100% であっても「とても正確であった」と評価するユーザは少ない．これは，字幕の精度以外の遅延や表示方法に何らかの不満があったことを示唆している可能性がある．ただし，単に精度が良かったと認めることに抵抗があっただけの可能性もあるため，さらなる検証が必要である．

今回の実験では TOEIC の試験問題を用いているため，実際の会議とは状況がかなり異なる．特に実際の会議は一方的な聞き取りではなく，双方向の対話であるため複雑である．分からない点はその場で聞き返せるため有利

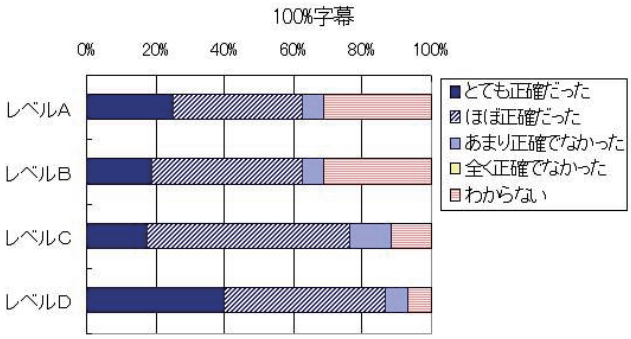

図 -4　字幕の正確さ

である一方，聞くだけではなく発言もしなければならないため，聞くことだけに集中するわけにもいかない．

また，英語字幕の別の効果として，英語の発話者側にも自分の発言の認識結果を表示することで，自分の発言の不明瞭さを気付かせ，明瞭に話すことを促す効果が期待できる．

このような効果を測定するためのシステムと実験手法の開発を現在行っている．

参考文献

1）Takezawa, T., Morimoto, T., Sagisaka, Y., Campbell, N., Iida, H., Sugaya, F., Yokoo, A. and Yamamoto, S.：A Japanese-to-English Speech Translation System：ATRMATRIX, Proc. ICSLP 1998, pp.2779-2782 (1998).
2）知野哲朗ほか：日中英 3 言語 6 方向音声翻訳システム，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).
3）Garza, T.：Evaluating the Use of Captioned Video Materials in Advanced Foreign Language Learning, Foreign Language Annals, Vol.24, No.3, pp.239-258 (1991).
4）Educational Testing Service：TOEIC PROFCIENCY SCALE.（オンライン）http://www.toeic.or.jp/toeic/pdf/data/proficiency.pdf
5）Educational Testing Service：TOEIC テスト新公式問題集，Vol.3 (2008).

（平成 21 年 11 月 4 日受付）

下郡信宏（正会員）
nobuhiro.shimogoori@toshiba.co.jp
1967 年千葉県生まれ．1992 年東京工業大学大学院修士課程修了．（株）東芝研究開発センターで知識処理の研究に従事．