

## 英語字幕による会議支援： 字幕の精度と表示タイミングが理解に及ぼす影響

下 郡 信 宏<sup>†1</sup> 池 田 朋 男<sup>†1</sup> 関 矢 陽 子<sup>†2</sup>

英語での会議を支援するために、英語の発言から音声認識で英語字幕を生成するシステムを検討している。TOEIC のレベル C の被験者 23 人で精度と表示タイミングの異なる字幕を表示しながら、TOEIC の長文リスニング試験をおこなった。字幕の精度は以前の実験で有効と確認された 80% と 100%、表示タイミングは発話前、発話 2 秒後、発話後の三種類を用いた。実験の結果、字幕の精度を 80% から 100% に向上させるよりも字幕を発話前に表示する方が理解に与える影響が大きいことを確認した。

### How display timing of captions affects comprehension of EFL speakers

NOBUHIRO SHIMOGORI,<sup>†1</sup> TOMOO IKEDA<sup>†1</sup>  
and YOKO SEKIYA<sup>†2</sup>

A system which assists English as a foreign language (EFL) speakers to attend meetings held in English by automatically generating English captions using automatic speech recognition is being studied. This paper describes how the display timing of the caption affects the comprehension of EFL speakers. We have conducted an experiment with 23 subjects, each with TOEIC level C English skills. The subjects took the TOEIC listening comprehension test while displaying captions with various timings and accuracy. Three types of caption display timing and two accuracy was tested. The results show that displaying the caption prior to each sentence is spoken contributed to the comprehension score rather than increasing the accuracy from 80% to 100%.

<sup>†1</sup> 株式会社 東芝 研究開発センター  
R&D Center Toshiba Corporation

<sup>†2</sup> 株式会社 東芝 CS 評価センター  
CS Evaluation Center Toshiba Corporation

### 1. はじめに

英語を母国語としない人が英語で会議をおこなうには多大な苦勞を伴う。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い場合もあり、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中、英語を母国語としない人が英語で会議をおこな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おこなわれている<sup>1)2)</sup>。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、機械翻訳による誤訳も含めて、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<sup>3)</sup>。

例えば、”The accuracy is pretty good.”という文の音声認識を失敗して”The accuracy is pretty goo.”と認識してしまった場合、goo は good の誤認識なのだと推測するのは比較的容易と思われるが、誤認識した文を翻訳すると「精度はきれいなべとべとしたものです」となってしまう、元の英文を推測する困難さは増す。

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、誤認識しても元の英文が推測しやすい。

更に、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。

音声認識により自動生成した字幕は、二つの点でテレビや映画の字幕と異なる。一点目は誤認識である。音声認識の認識精度には限界があり、講義や会議における発話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<sup>4)</sup>。そこで筆者らは音声認識に求められる精度を測定する実験をおこない、精度が 80% 以上であれば日本人の約半数に有効であることを実験により確認して報告した<sup>5)</sup>。二点目は字幕の表示タイミングである。一般にテレビや映画の字幕は発言の直前に表示されるが、音声認識を用いて作成する場合は、発言が終わるまで字幕が確定しないため、発言後にしか表示できない。そこで、今回は字幕の表示タイミングが理解に及ぼす影響を実験により測定した。

## 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。Graza 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<sup>6)</sup>。また、Huang 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<sup>7)</sup>。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かった。そこで筆者らは音声認識により生成した字幕を英語を第二外国語とする被験者に見せて、単語正解精度が 80% の字幕であってもリスニングの理解度が向上することを確認した。

一方で、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。

テレビ放送でも特にニュースは生放送であるため字幕のずれが問題になる。そこでニュース字幕の表示タイミングのずれに関する研究が行われている。金澤らはテレビ放送のニュース字幕の表示タイミングのずれは映像よりも字幕が約 2 秒遅い条件から約 4 秒早い条件までが許容の範囲であるとした<sup>8)</sup>。一方丸山らは聴覚障害者にとって許容されるニュース字幕のずれを測定しており、その中で比較のための健常者グループの許容範囲が約 ± 2 秒であることを明らかにした<sup>9)</sup>。しかしこれらの実験は被験者の好みの主観評価であり、字幕のずれによる理解度の変化を定量的に測定したものではない。

筆者らが先におこなった実験では発話直後に字幕を表示していたため、発話直後も有効である事は確認されている。しかし、精度とタイミングを分離して測定していないため、タイミングを変更すると理解度が向上するか明らかでない。そこで、字幕の表示タイミングが内容の理解に影響を与えるか確認する必要があった。

## 3. 実験

以下の仮説を検証するための実験をおこなった。

仮説 1 字幕を発話前に表示した方が理解度が高い

発話前に字幕を表示すると、音声を聞きながら同時に字幕を追うことが可能になるため理解しやすいと考えられる

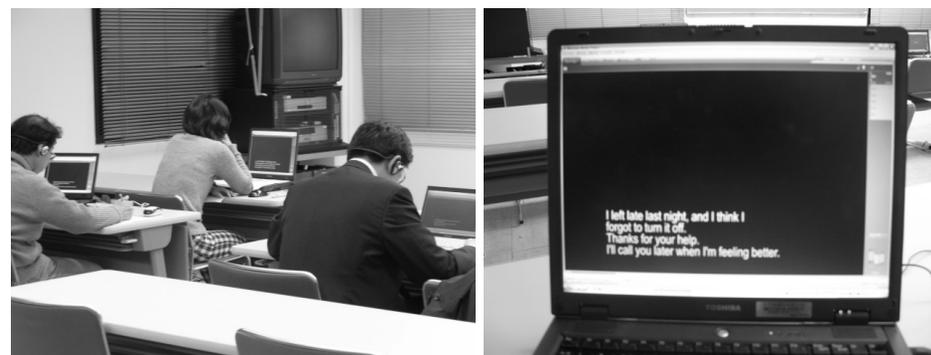


図 1 実験風景

Fig. 1 Experiment Environment

図 2 字幕表示画面

Fig. 2 Caption Display Screen

仮説 2 字幕の表示タイミングが早ければ精度が悪くても理解度は上がる

発話前に字幕を表示すれば、字幕の間違いに気づきやすいため、精度を上げるよりもタイミングを早める方が成績の向上につながりやすいと考えられる

### 3.1 被験者

23 人の被験者を用いて実験をおこなった。筆者が行った要求される精度を求める実験で TOEIC のレベル C の人に英語字幕の効果が確認されたため、今回の被験者は TOEIC レベル C の人に限定した。被験者はいずれも日本人の社会人で TOEIC レベル C の成績を 2 年以内に取得している。TOEIC のレベル C とは 470 点 ~ 730 点の得点を有する人をさし、「日常生活のニーズを充足し、限定された範囲内では業務上のコミュニケーションができる」<sup>10)</sup> とされるレベルであり、日本人の受験者の約半数がこのレベルに該当する。その他の被験者のプロフィールは以下の通り。TOEIC のレベルを制限した以外は被験者のプロフィールによる選別を行っていない。ただし、実験で使用される TOEIC の試験問題は市販の問題集であるため、いずれの問題集も使用していないことを被験者の募集時に確認してある。被験者には実験の参加に対して報酬が支払われている。

(性別) 男性: 13 名, 女性 10 名

(年齢) 20 代: 13 名, 30 代: 5 名, 40 代: 5 名

(得点) 470 ~ 530 4 名, ~ 580 3 名, ~ 630 4 名, ~ 680 7 名, ~ 730 5 名

### 3.2 試験問題

試験問題には TOEIC のリスニング試験の Part 4 を用いた。Part 4 は 100 語前後の長文が読み上げられ、その文に関連する質問が 3 問出題される形式の問題である。読み上げられる長文は音声でのみ出題され、手元の問題用紙には記載されていない。問題文と解答の選択肢は問題用紙に記載されている。読み上げ文と問題三問が一つのセットになっており、通常の TOEIC の試験には 10 セット含まれている。今回の実験では字幕の設定毎に試験 1 回分 (10 セット) の問題を使用し、合計 7 回分の試験をおこなった。これは各セット間の難易度にはバラツキがあるが、10 セット集まると平均化され、試験間での難易度のバラツキが小さくなるためである。試験問題は TOEIC の公式問題集を用いた<sup>11)12)13)14)</sup>。

### 3.3 字幕

字幕の精度は 80% と 100% の二種類を用意し、それを発話前、発話開始 2 秒後、発話後の三つのタイミングで表示した。発話前ではこれから発話する一文を発話が始まる直前に一気に表示する。これは発話と字幕のズレがない状態であり、テレビでもドラマなどの録画された映像ではこの表示方式が多く採用されている。発話開始 2 秒後では発話が始まった 2 秒後に 2 秒間の間の発言内容と共に、発話中の文の最後までを一気に表示する。丸山らが字幕のずれの許容範囲を ± 2 秒としたことからこの値を採用した。また、金澤らの + 4 秒にすると次の発話後との重なりが多くなるため 2 秒とした。発話後では一文の発話が完了してから、その内容を一気に表示する。この発話後の表示方法は音声認識を用いて字幕を表示した場合の表示方法と同じである。いずれの場合も、一文が長く、表示領域の 4 行に収まらない場合は、発話の途中の息継ぎまでを一文とした。最後に比較のために字幕を表示しない通常の TOEIC 試験と同じ状況での試験もおこなったため、以下の合計 7 種類の字幕環境で実験をおこなった。

- (1) 精度 100%，発話前
- (2) 精度 100%，発話開始 2 秒後
- (3) 精度 100%，発話後
- (4) 精度 80%，発話前
- (5) 精度 80%，発話開始 2 秒後
- (6) 精度 80%，発話後
- (7) 字幕なし

精度は音声認識の分野で一般的に用いられている単語正解精度を用いており、以下の計算式で求められる。

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

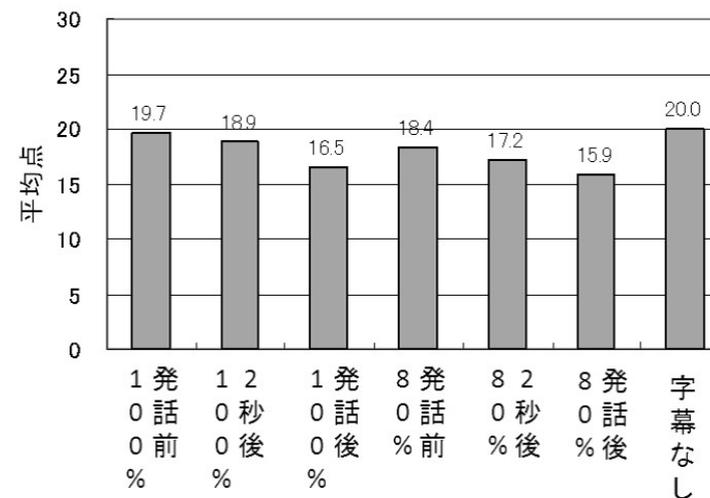


図 3 成績  
Fig. 3 Score

また、音声認識特有の誤りを再現するため、試験の読み上げ文を音声認識ソフトで認識させて誤認識を発生させた。音声認識ソフトには Dragon NaturallySpeaking 10 をエンロールなしで使用している<sup>15)</sup>。問題文によっては 80% 以上の精度で認識できてしまう。そのため、80% 以上の精度で認識できたしまった問題に関しては、音声データにノイズを混ぜて認識精度を落として、80% 以下の精度の字幕を生成した後に、精度が 80% になるまで誤認識箇所を正解に戻して行く作業をおこなった。80% 以上の精度で認識できた問題では 80% 以上の認識結果で誤認識している箇所を残す様にして、より実際の認識結果に近づけた。更に、正解に戻す際には、誤認識が一箇所に固まらないように分散させた。

字幕は一行 40 文字で画面の下 1/3 に 4 行分を表示した。新しい行が追加されると古い行は上にスクロールし、常に 4 行表示されている状態を保つ。複数行を一気にスクロールする時には先頭行が分からなくなるように、一度に複数行追加せずに、一行ずつ素早く複数回追加することでスクロールした量を把握できる様にした。字幕以外には黒い背景のみが表示されている。

### 3.4 実験手順

複数人の被験者が同じ部屋で試験を受けた。問題や字幕の順番による影響を排除するため、被

験者に与える問題の順番，字幕の種類と問題の組み合わせ，さらに字幕の試験順序も被験者毎に異なるようにした．複数の被験者が同じ部屋で異なる試験問題を受験するため，各被験者はヘッドホンを用いて各自のパソコンに字幕を表示しながら試験を受けた．

被験者は最初に4セット12問の練習問題を行う．この練習問題で字幕が表示される全てのタイミングを経験しながらヘッドホンの音量の調整を行う．被験者は全員 TOEIC を受験しているので，問題の解き方などは既に習熟している．

練習問題で分からない点などが無いことを確認した後に，試験を開始する．合計7回分の試験のうち，3回分が終わった段階で5分の休憩を取り，引き続き4回分の試験を行う．試験には合計2時間を要した．7回の試験が1回終わる度に，官能評価のためのアンケートを行っている．質問項目は字幕のタイミングの満足度と字幕の正確さの満足度をそれぞれ5段階で評価する．

#### 4. 結 果

図3は字幕の種類別の正解数の平均である．各30問あるため，30点で全問正解である．字幕の精度が同じグループでは表示タイミングが早いほど平均点が高い．ただし，字幕なしの実験条件の平均点が最も高くなっており，字幕を表示しない方が理解度が上がる結果になっている．精度80%の字幕で効果があるはずなので，これは期待と異なる．これに関しては考察で述べる．

##### 4.1 仮説の検証

最初に「(仮説1)字幕を発話前に表示した方が理解度が高い」を検証する．

字幕ありの中で，正解数にどの要因が効果があるかを確認するため，最小二乗法による分析をおこなった．要因を人，提示順，字幕，問題とし効果の検定をおこなったところ，「提示順」以外の要因は有意であった(図4)．そこで提示順を除いた3要因で分析をおこなった(図5)．

更に「字幕」の最小二乗平均について多重比較検定を行うと，図6の様になる．

その結果，有意水準5%以下に有意差があった．有意差のあるセルは下線で表している．

- 100% 発話前と100% 発話後
- 100% 発話前と80% 発話後
- 100%2秒後と100% 発話後
- 100%2秒後と80% 発話後

これにより，発話後に表示される字幕は常に劣っていることがわかる．ただし，80% 発話前と80% 発話後，100% 発話後のいずれの間にも有意な差は見られなかった．このため，仮説1は部分的にしか確認することができなかった．

次に「(仮説2)字幕の表示タイミングが早ければ精度が悪くても理解度は上がる」を検証す

あてはめの要約				
R2乗				0.663848
自由度調整R2乗				0.530073
誤差の標準偏差(RMSE)				2.944198
Yの平均				17.76087
オブザベーション(または重みの合計)				138

分散分析				
要因	自由度	平方和	平均平方	F値
モデル	39	1677.6149	43.0158	4.9624
誤差	98	849.4998	8.6683	p値(Prob>F)
全体(修正済み)	137	2527.1087		<.0001*

効果の検定					
要因	パラメータ数	自由度	平方和	F値	p値(Prob>F)
提示順	6	6	65.3564	1.2566	0.2846
字幕	5	5	233.5583	5.3888	0.0002*
問題	6	6	261.0890	5.0200	0.0002*
人	22	22	1101.2367	5.7746	<.0001*

図4 最小二乗法による分析(全要因)  
Fig.4 Least-Squares Analysis (All elements)

あてはめの要約				
R2乗				0.637995
自由度調整R2乗				0.523115
誤差の標準偏差(RMSE)				2.965912
Yの平均				17.76087
オブザベーション(または重みの合計)				138

分散分析				
要因	自由度	平方和	平均平方	F値
モデル	33	1612.2585	48.8563	5.5540
誤差	104	914.8502	8.7966	p値(Prob>F)
全体(修正済み)	137	2527.1087		<.0001*

効果の検定					
要因	パラメータ数	自由度	平方和	F値	p値(Prob>F)
字幕	5	5	237.7016	5.4044	0.0002*
問題	6	6	253.6571	4.8059	0.0002*
人	22	22	1108.2389	5.7266	<.0001*

図5 最小二乗法による分析(提示順除く)  
Fig.5 Least-Squares Analysis (caption order excluded)

最小二乗平均差のTukeyのHSD検定						
α = 0.050 Q = 2.90355						
平均(±平均)の標準誤差	最小二乗平均					
	100% 発話前	100% 2秒後	100% 発話後	80% 発話前	80% 2秒後	80% 発話後
100% 発話前	0	0.56687	3.18427	1.15815	2.35899	3.22312
	0	0.87661	0.87668	0.8787	0.87852	0.88028
	0	-1.9784	0.63279	-1.3932	-0.1918	0.65724
	0	3.11216	5.72975	3.7095	4.90981	6.08051
100% 2秒後	-0.56689	0	2.61739	0.59128	1.79212	2.95625
	0.87661	0	0.87852	0.87846	0.87879	0.87762
	-3.1122	0	0.06558	-1.9594	-0.7595	0.40304
	1.97842	0	5.16821	3.14194	4.34374	5.50446
100% 発話後	-3.1843	-2.6174	0	-2.0261	-0.8253	0.3886
	0.87668	0.87852	0	0.87854	0.88365	0.88247
	-5.7297	-5.1682	0	-4.577	-3.391	-2.2234
	-0.5668	-0.0656	0	0.52476	1.74045	2.90114
80% 発話前	-1.1582	-0.5913	2.00511	0	1.20384	2.35497
	0.8787	0.87846	0.87854	0	0.88266	0.87849
	-3.7095	-3.1419	-0.5248	0	-1.362	-0.1858
	1.3932	1.95938	4.57699	0	3.76367	4.91571
80% 2秒後	-2.359	-1.7921	0.82928	-1.2008	0	1.16413
	0.87852	0.87879	0.88365	0.88266	0	0.88389
	-4.9098	-4.2437	-1.7405	-3.7637	0	-1.4023
	0.19184	0.75951	3.39101	1.362	0	3.73056
80% 発話後	-3.2231	-2.9563	-0.389	-2.365	-1.1641	0
	0.88078	0.87762	0.88247	0.87849	0.88389	0
	-6.0805	-5.5046	-2.9011	-4.9157	-3.7306	0
	-0.8657	-0.4030	2.22343	0.18577	1.40229	0

水準	最小二乗平均
100%発話前 A	19.591153
100%2秒後 A	19.024282
100%発話後 A B	18.433002
80%発話前 A B	17.232164
80%2秒後 B	16.406887
80%発話後 B	16.068030

AまたはBでつながっていない水準は有意に異なる

図6 最小二乗平均差のTukeyのHSD検定  
Fig.6 LSMeans Differences Tukey HSD

あてはめの要約	
R2乗	0.633225
自由度調整R2乗	0.52596
誤差の標準偏差(RMSE)	2.957053
Yの平均	17.76087
オブザベーション(または重みの合計)	138

分散分析				
要因	自由度	平方和	平均平方	F値
モデル	31	1600.2275	51.6202	5.9034
誤差	106	926.8812	8.7442	p値(Prob>F)
全体(修正済み)	137	2527.1087		<.0001*

効果の検定					
要因	パラメータ数	自由度	平方和	F値	p値(Prob>F)
精度	1	1	41.2299	4.7151	0.0321*
タイミング	2	2	184.2682	10.5366	<.0001*
問題	6	6	248.2492	4.7317	0.0003*
人	22	22	1108.4225	5.7619	<.0001*

図7 最小二乗法による分析(精度とタイミング)

Fig. 7 Least-Squares Analysis (accuracy and timing)

る。字幕精度 80% 発話前の方が字幕精度 100% 発話後よりも平均点は高いが、先の分析で有意な差は確認できていない。そこで、「精度」と「タイミング」の影響を見るために、「字幕」としていった要因を「精度」と「タイミング」に分解して効果の検定を再度行った(図7)。この結果、タイミングのF値と精度のF値を比較すると、タイミングの方が正解数に及ぼす影響が大きいことが分かる。従って、タイミングが早ければ精度が悪くても理解度は上がるとまでは言い切れないものの、タイミングの方が精度よりも理解度に大きな影響を及ぼすと言うことができる。

#### 4.2 アンケート結果

試験の各回が終了する度に行ったアンケートの結果は以下の通りである。図8では字幕が正確であったと感じたかどうかを聞いている。精度100%の字幕を発話前に表示した場合は、全員がとても正確、またはほぼ正確としており正しく判定出来ている。タイミングが遅れるに従って、とても正確、ほぼ正確とした人が減少していき、正しく判定できていないことがわかる。一方精度80%の字幕では、発話前に表示した場合はあまり正確でない、全く正確でないが多く、タイミングが遅くなるに連れ字幕が正確、または分からないとする人の数が増える。こちらでも正しく判定できていないことが確認できる。

次に、字幕表示のタイミングの満足度の結果を説明する。図9がその結果である。100% 発話前が一番満足度が高く、続いて80% 発話前の満足度が高いことが特徴的である。逆に発話後は精度100%と精度80%のどちらでも満足度が低く表示タイミングの重要性を示している。

最後に、アンケートの自由記述欄に記載された代表的なコメントを記載する。

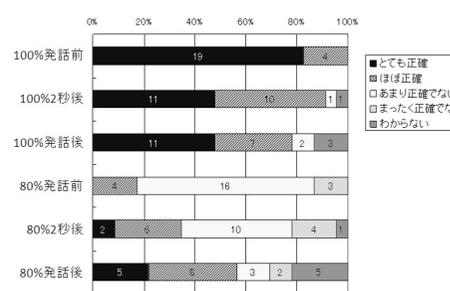


図8 字幕は正確でしたか?

Fig. 8 Was the caption accurate?

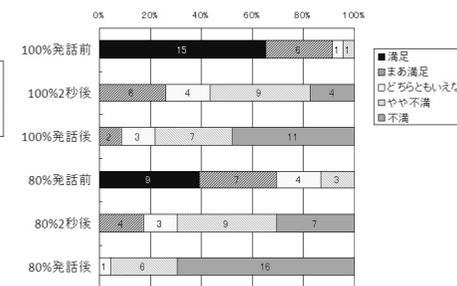


図9 タイミングは満足しましたか?

Fig. 9 Where you satisfied with the display timing?

- 字幕を見なければならぬと問題文を読む暇がなくなり、結局聞き取れない
  - TOEIC としては問題を見ながら聞くというほうが大事な部分を聞き取れるのでいい
  - 字幕の出方が非常に見づらくて、次の文章がどこから始まるか目で追っていくことが困難
  - 字幕のスピードが発音と同時に聞こえないので字幕を見て回答する方法はとても難しく感じた。字幕の出るタイミングがもっと速ければ発音と同時に理解できると思う。
  - 字幕が間違っていると気をとられてその後の聞き取りが難しかった。
- これらの指摘は特に複数人の被験者から同様の指摘があったものである。

#### 5. 考 察

まず、字幕なしの方が100%の字幕を発話前に表示した場合よりも点数が高かった現象に関して考察する。最後のアンケートでも紹介した通り、問題文を読む時間がないと多くの被験者がコメントしている。TOEICのPart 3と4は長文ヒアリング問題である。1分程度の間に読み上げられた情報を理解できたとしても、具体的な人名や時刻などは予め意識していないと憶え切れないことがあり、それが原因で点数が下がることがある。これを克服するために、問題文の先読みと呼ばれる試験テクニックが広く知られている。読み上げ文を聞く前に問題文を読むことで、聞き取る対象を絞り込むテクニックである。例えば問題文の先頭がWhoで始まっていたら人物、Whenなら日時、Whereなら場所に集中して聞き取るようにしたり、解答の選択肢の中に時刻や数字が列挙されていたら、それらに注意するなど、問題文を全部読む事なく絞り込みを行う。今回の実験では字幕がある時は字幕を見るように指示したため、問題文の先読みのテクニックが使えず、結果として先読みができた字幕なしの点数が高くなったと考えられる。字幕あ

りと字幕なしの実験条件を揃えるためには、どちらでも問題文の先読みが出来る様に十分な時間を与えるか、問題文の先読みができない様に問題用紙を配らず、解答するタイミングで画面に問題を表示するなどの工夫が必要であった。

次に字幕のタイミングに関して考える。発話前に字幕を表示すると耳で聞き取りながら字幕を追いかけることが可能であるため、被験者の負荷が少ない。また、字幕に間違いがある場合も字幕に間違いがあることが判定しやすいため混乱も少ない。一方で発話後に字幕を表示すると耳で一旦理解した内容を記憶した後に表示されたテキストと照合しなければならない。更に、照合中も次の音声聞き取らなければならないため、並列して処理を行う必要があり負荷が増大する。上記の様に行っている作業が本質的に異なるために、正解数からもアンケート結果からも差が出るものと考えられる。

最後に字幕の表示方法について述べる。字幕の表示方法に関する不満が多く聞かれた。字幕の先頭を見失わない様に複数行を表示する場合は一気に切り替えずに、一行ずつ複数回に分けて素早く書き換えたが、それでも先頭を見失うことがあった模様だ。テレビや映画などの字幕では発言毎に全ての字幕を消して次の字幕を表示するのが一般的なため、常に先頭から読めば良いようになっている。しかし、今回の実験でこの方式を採用すると、長い発言の後に短い発言があった場合、長い発言が終了してから、長い字幕が表示され、長い字幕を読んでいる最中に次の短い発言が終わってしまい、字幕を読み終わる前に消えてしまう問題が発生する。そこで、4行文の字幕は残した。このような表示方法を行う場合は、新しく表示された字幕と古い字幕で文字の色を変えるなど、更に工夫が必要であった。

## 6. おわりに

音声認識で生成された誤りを含む字幕の表示タイミングの影響を測定する実験を行った。その結果、発話前に字幕を表示する方が字幕の精度を向上させるよりも理解度への影響が大きいことが分かった。また、アンケートからも、発話前に字幕を表示した方が被験者は字幕の誤りを正しく識別出来ていることが分かった。対面の会議で話者の発言前に字幕を表示させることは予め発言内容が決まっている場合を除き技術的に不可能だが、遠隔会議であれば数秒の音声の遅延が許容される場面がある。音声認識で字幕を生成して会議を支援するようなシステムでは、音声認識の精度だけでなく、字幕の表示タイミングも重要であることが分かった。今後は、実際のシステムで効果を検証する予定である。

## 参 考 文 献

- 1) T. Takezawa, T. Morimoto, Y. Sagisaka, et al: A Japanese-to-English speech translation system: ATR-MATRIX, Proc. ICSLP 1998, pp.2779-2782(1998).
- 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, 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) 菅谷史昭, 竹澤寿幸, 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, Vol.J84-D-II, No.11, pp.2362-2370(2001).
- 4) Whittaker, S. and Hirschberg, J.: Look or Listen: Discovering Effective Techniques for Accessing Speech Data. Proc. *Human-Computer Interaction Conference 2003*, pp.253-269.(2003)
- 5) 下郡信宏, 坪井創吾: 会議における英語字幕の効果, マルチメディア, 分散, 協調とモバイル (DICOMO2009) シンポジウム論文集, pp687-694(2009).
- 6) Garza, T.: Evaluating the Use of Captioned Video Materials in Advanced Foreign Language Learning, *Foreign Language Annals*, Vol.24, No.3, pp239-258(1991).
- 7) Huang, H. and Eskey, D.: The effects of closed-captioned television on the listening comprehension of intermediate English as a second language (ESL) students, *J. Educational Technology Systems*, Vol.28, No.1, pp75-96(2000).
- 8) 金澤章, 磯野春雄: ニュース字幕の提示タイミングずれの主観評価と補正方法, 映像情報メディア学会年次大会講演予稿集, No.2001, pp89-90(2001).
- 9) 丸山一郎, 阿部芳春, 沢村英治ほか: ニュース字幕の提示タイミングずれに対する許容特性, 電子情報通信学会技術研究報告. HCS, ヒューマンコミュニケーション基礎”, Vol.99, No.123, pp21-28(1999).
- 10) Educational Testing Service: TOEIC PROFICIENCY SCALE. (オンライン), 入手先 <<http://www.toeic.or.jp/toeic/pdf/data/proficiency.pdf>> (参照 2010-02-22)
- 11) Educational Testing Service: TOEIC テスト新公式問題集 <Vol. 1>, 国際ビジネスコミュニケーション協会 (2006).
- 12) Educational Testing Service: TOEIC テスト新公式問題集 <Vol. 2>, 国際ビジネスコミュニケーション協会 (2007).
- 13) 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).
- 14) Educational Testing Service: TOEIC テスト新公式問題集 <Vol. 4>, 国際ビジネスコミュニケーション協会 (2009).
- 15) Nuance Communications, Inc. :Dragon NaturallySpeaking (オンライン), 入手先 <<http://www.nuance.com/naturallyspeaking/>> (参照 2010-02-22)