様　式　Ｃ－１９、Ｆ－１９、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

平成 ２７ 年　 ６ 月 ２２ 日現在

機関番号：１２１０３
研究種目：基盤研究(C)
研究期間：2011～2014
課題番号：２３５０１０９６

研究課題名（和文）聴覚障害者のための字幕システムにおける字幕とノンバーバル情報の呈示法に関する研究

研究課題名（英文）A study for displaying timing between verbal and non-verbal information for a real-time speech-to-caption system

研究代表者

黒木　速人（KUROKI, Hayato）

筑波技術大学・産業技術学部・教授

研究者番号：００３４５１５９

交付決定額（研究期間全体）：（直接経費）　　3,800,000円

研究成果の概要（和文）：研究代表者らは，聴覚障害者のための復唱音声認識方式によるリアルタイム音声-字幕変換システムを開発している．本研究では，字幕呈示過程において，限られた精度の字幕に対し話者発話時のノンバーバル情報を最適に呈示させることで最終的な内容理解を高めることを目的とする．
実験では，話者発話時の顔映像と，音声認識誤認識文の呈示タイミングを操作した試料から，被験者がどれだけ正確に読み上げ文を回答できるかを比較した．
結果，字幕先行呈示の領域において，字幕先行である程正答率が上昇し，顔映像に対して字幕が1秒先行する条件では，聴覚障害者群においては正答率が最大値を示し，健聴者群においては最大値と同等の値を示した．

研究成果の概要（英文）：This study aimed to investigate the suitable ways of the display display sequence and display timing between captions that have speech recognition errors and the speaker's face movement images for our developing real-time speech-to-caption system for the hearing-impaired.
The results showed that the sequence displaying the caption before the speaker's face image improved the comprehension of the captions. The sequence displaying both simultaneously showed an improvement of only a few percent higher than that of the question sentence, and
the sequence displaying the speaker's face image before the caption showed almost no change. In addition, the sequence displaying the caption 1 second before the speaker's face showed the most significant improvement of all the conditions in the hearing-impaired.

研究分野：アシスティブテクノロジー

キーワード：ヒューマン・インタフェース　障害者支援　高等教育支援　教育工学　認知科学

１．研究開始当初の背景

聴覚障害者に対する情報保証手段として，音声認識技術の持つ利点である「音声入力による簡便さ」と「文字化までのリアルタイム性の高さ」を積極的に活用する試みがなされている．しかし音声認識処理に関する現在の技術レベルでは，話し手が交代した場合の不特定話者の認識や，話し言葉のような厳格ではない文法構造に対する認識，未知語の認識などは技術的に非常に難しく，音声認識技術を実用的な情報保障手段として用いるにはまだまだ技術的な工夫が必要である．

一方，ヒトは，自由発話（話し言葉）に良く見られるような語の省略，倒置，不要語が追加された文や，多少の誤りや未知語が含まれる文に対しても，文脈，単語の前後関係やノンバーバル情報（話し手の，表情，口の動き，ジェスチャー等）などから意味を類推して正しい内容を理解することができる．これには，曖昧な文でもそれを理解するヒトの認知能力の関与するところが大きい．このような背景のもと，情報保障手段として音声を文字化させることの有効性に着目し，音声認識技術の持つ利点とヒトの持つ認知特性を生かして，音声-字幕変換を行う新しいタイプのシステムを研究・開発している．現在までに，システムの字幕生成過程における設計として，音声認識装置の弱点である不特定話者認識，話し言葉認識，未知語処理を克服するために，復唱者による音声認識方式と字幕修正者による認識結果修正方式を採用し，対遠隔地運用試験を実施してきた．

２．研究の目的

本研究では，本システムの字幕呈示過程におけるシステムの設計として，限られた精度の字幕から，より高い内容理解を得るための字幕とノンバーバル情報の呈示方法に関する研究を行う．具体的には，情報取得者が最終的に獲得する内容理解を向上させるために，話者の発話時の「顔」と，音声認識処理の誤認識結果である「誤り」を含み，文としては「不完全な文」を，どのような順序，時差で呈示する方法が，内容理解の促進あるいは阻害に繋がるかを定量的に把握する．加えて，ヒトが参照する情報は，適切な形で呈示されないと逆にヒトの内容理解を阻害する恐れがあるため，情報の呈示に関しても適切な方法を探求していくことが必要となる．

３．研究の方法

実験では，話者の発話時の顔映像と，誤認識を含む不完全文の呈示タイミングを操作した試料をいくつか用意し，被験者がどれだけ正確に元の完全文（音声認識処理を施す前の原文，正解文）を回答できるかを比較した．ここで，顔映像と字幕の呈示タイミングとは，二つの情報の呈示順序・呈示時差のことを示す．原文に対する回答文の形態素レベルでの正解率を回答文完全率として算出し，内容理解の促進・阻害の度合を測る指標とした．

原文として用いた文は，日本音響学会編「研究用連続音声データベース」の音素バランス文から引用した．これらの原文に対し，音声認識処理を施すことで不完全文とした．音声認識処理には，認識単語の確からしさの一尺度として「尤度」があり，尤度は認識結果の各々の形態素に付帯してくる．尤度閾値を設定することで，閾値以下の認識結果を「＊」で置換することができる．つまり不完全文は，正しい認識結果，誤った認識結果（誤認識結果），尤度閾値以下の結果（「＊」で置換）の 3 種類の文字から構成されることになる．

作成した試料の呈示の順序は，時差±0 秒，－1 秒（字幕先行呈示 1 秒），－2 秒（字幕先行呈示 2 秒），…，－5 秒，－5 秒，…，－1 秒，±0 秒，+1 秒（顔先行表示 1 秒），+2 秒，…，+5 秒，+5 秒，…，+1 秒，…，の繰り返しとし，各時差に付き 10 題，合計 110 題の呈示試料を事前録画したものを用いた．学習効果が入らないようにするために一度使用した原文の重複使用は避けた．

４．研究成果

原文に対する回答文の形態素レベルでの正解率を回答文完全率として，情報の呈示時差ごとに，各被験者に対する結果と被験者群ごとに算出した．

(1)被験者ごとの結果

聴覚障害者群においては，どの被験者も類似した傾向を示した．字幕先行呈示の領域においては，呈示時差の値が－2 秒以下となるほど（呈示時差軸において－2 秒よりも左側（－側）に値が行くほど），各被験者の回答文完全率の値は上昇する傾向が認められた．呈示時差－1 秒（顔映像に対し字幕を 1 秒先行呈示）においては，全呈示時差を通して最も高い回答文完全率を示した．

呈示時差±0 秒（顔映像と字幕の同時呈示）は同期呈示であり，日常的には最も自然な呈示方法であるため最も高い値を示すと予想されたが，結果は課題文完全率よりも数%上昇するに留まった．

顔先行呈示（呈示時差軸において右側（+側））の領域においては，顕著な傾向は示さなかった．

健聴者群においては，呈示時差－1 秒に極値（全域的な極大でなくとも極値を取る）を持つ傾向，顔先行呈示の領域においては顕著な傾向は確認できないなど，聴覚障害者群と

似たような傾向がいくつか認められた．呈示時差－1 秒における値が極大値でなく極値であるなど，傾向は顕著なものとしては認められなかった．また，健聴者群は聴覚障害者群と比較して個人差が大きいことが見て取れた．

(2)被験者群平均の結果

聴覚障害者群と健聴者群に被験者群を分けて解析した結果を図 1 と図 2 に示す．

聴覚障害者群においては，字幕先行呈示領域において，呈示時差－5 秒（$p<0.05$），呈示時差－4 秒（$p<0.01$）と呈示時差－1 秒（$p<0.01$）において有意差を示した．また，顔先行呈示領域においては，呈示時差+3 秒（$p<0.01$）において有意差を示した（図 1）．

健聴者群においては，呈示時差－5 秒（$p<0.05$），－4 秒（$p<0.01$）と－1 秒（$p<0.01$）において有意差を示した（図 2）．

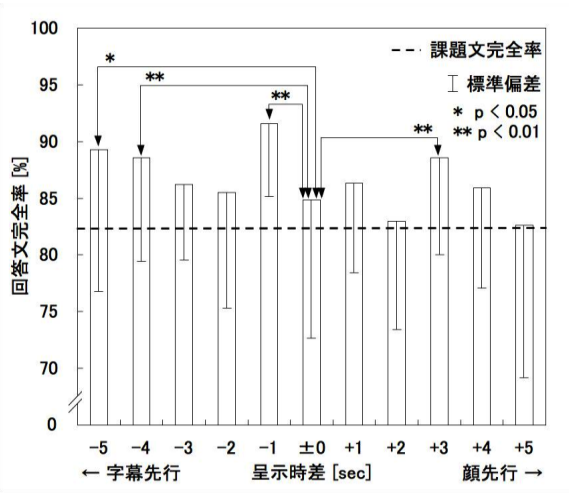

図 1 聴覚障害者群

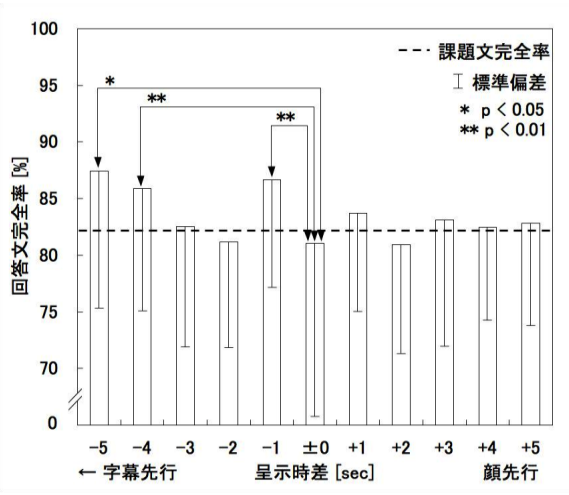

図 2 健聴者群

(3)まとめ

字幕先行呈示（呈示時差軸の左側（負の呈示時差)）領域において，呈示時差の値が－2 秒以下となるほど回答文完全率が上昇する傾向が認められた．呈示時差－1 秒（顔映像に対して字幕が 1 秒先行呈示）に関しては，聴覚障害者群においては，回答文完全率が全域的な極大値を示し，健聴者群においては，全域的な極大値は呈示時差－5 秒で示されたが，呈示時差－1 秒では極大値とほぼ同じ値を示した．この理由として以下のことが考えられる．字幕は時間とともに次々と呈示内容が更新されるが，ある一定時間内において課題文はディスプレイ上に「字幕」として留まる．そのため被験者は，消失するまでの一定時間内は課題文自体や文中の誤認識を何度も読み返すことが可能になる．さらに，停留する字幕に対し，顔映像をある一定の呈示時差を設けて呈示することで，字幕と顔映像と言った二つの異なるモダリティから情報を読取るタスクにおいて，効果的な呈示タイミングになったのではないかと推察される．とりわけ呈示時差－1 秒は非常に効果が高く，文字通りタイミングの合った呈示方法であると言える．この傾向は特に聴覚障害者群で顕著に認められ，健聴者群においても同様の傾向を示した．

５．主な発表論文等

（研究代表者、研究分担者及び連携研究者には下線）

〔雑誌論文〕（計 1 件）

①黒木速人，井野秀一，中野聡子，堀耕太郎，伊福部達，阿山みよし，長谷川光司，湯山一郎，聴覚障害者のためのリアルタイム字幕システムにおける話者顔映像と誤認識字幕の呈示タイミングに関する研究，映像情報メディア学会論文誌，Vol.65, No.12, pp.1750-1757, 2011(査読あり).
DOI: 10.3169/itej.65.1750

〔その他〕

ホームページ等

筑波技術大学・機関リポジトリ

http://www.a.tsukuba-tech.ac.jp/repo/dspace/

６．研究組織

(1)研究代表者

黒木 速人（KUROKI, Hayato）
筑波技術大学・産業技術学部・教授
研究者番号：00345159

(2)研究分担者

なし

(3)連携研究者

伊福部 達（IFUKUBE, Tohru）
東京大学・高齢社会総合研究機構・名誉教授
研究者番号：70002102

中野 聡子（NAKANO, Satoko）
広島大学・アクセシビリティセンター・特任講師
研究者番号：20359665