

話し言葉のパラ言語情報を反映したテキスト構成手法の提案

内田 祥一 安井 浩之 吉野 邦生
東京都市大学

1. 目的と背景

近年、障害者支援が様々な分野において進められており、教育分野においては障害者に対する情報保障のニーズが高まっている[1]。聴覚障害者を支援する手法の1つとしてリアルタイム字幕呈示システムがあるが、音声認識から得られた文に含まれている誤認識文字を手動で修正する作業が必要であり、その後に聴覚障害者に字幕が提供されている[2]。テレビや映画などの様々な映像には、発話者の姿や表情が見えないものやナレーションがあるが、そのような状況では、たとえ映像に字幕が与えられていたとしても、音声で表現された発話者の感情を聴覚障害者が認知することは困難である。また、音声認識結果を正しく取得でき、誤認識文字がない場合であっても、話し言葉を忠実に文字化した字幕は、話し言葉固有の性質により読みにくさが残されてしまっている場合がある。

本報告は、話し言葉をテキスト化することによって生じる「パラ言語情報の欠落」に着目し、音声特徴量により得られた情報を基にパラ言語情報を特定し、映像の字幕情報をパラ言語情報が含まれたテキスト構成にすることで、話し言葉を字幕化の際の問題を解決し、聴覚障害者にとって読みやすい字幕呈示を行う手法を提案することである。

2. 関連研究

有賀ら [3]の研究では、先行研究によって定義されていた、教師の自発音声の特徴量から抽出できる6つのパラ言語情報「無関心」、「落胆」、「疑い」、「感心」、「中立」、「強調」に加えて、新たに「納得」、「気付かせ」、「反語」、「確認」の4つが、教師の自発音声からパラ言語情報として抽出され、定義された。

また、水口ら [4]は、文字の位置や大きさなどの属性を時間的に変化させることによって文字情報を映像的に表現する文字アニメーションを自動合成する試みを行っている。文字そのもの

のが持つ具体的な意味を、動きやタイポグラフィなどの視覚要素で暗示的に付与できる情報の視覚化ツールとしての可能性に言及している。

3. 研究内容

本研究は、字幕が挿入されている動画の字幕部分に対してパラ言語情報を付加することで聴覚障害者が動画を閲覧する際により動画内容を理解しやすい形にすることを目的とする。

具体的には、音声分析ソフトを用いて発話の音声特徴量を抽出し、パラ言語情報を特定する。パラ言語情報の特定に利用する特徴量は「平均基本周波数」、「最高基本周波数」、「最低基本周波数」、「ピッチレンジ」、「持続時間」の5つとする。その特定できたパラ言語情報を文字アニメーションで表現することにより、発話者が音声に込めた感情を伝えることを目指す。

なお、アニメーション化は文章すべてではなく、発話の中でも音声特徴量が特に変化する部分のみをアニメーション化する。

本研究では、YouTube[5]で公開されているナレーションと字幕が挿入されている教育用動画を対象とし、想定する利用環境は、聴覚障害を持つユーザが一人で教育用映像をPCで見ながら学習を行う際の補助である。

YouTubeには標準で字幕挿入機能があり、ユーザは自分が投稿した動画に字幕を付けることができる。字幕情報は、タイムコードと動画に挿入したい文字から作成されている。このテキストデータをYouTubeに投稿されている動画から抽出し、パラ言語情報を付加した新たな字幕を付与する。今回は字幕に付加するパラ言語情報に「強調」「感心」の2つを利用する。

「強調」のアニメーションには文字が一度拡大された後、元のフォントに戻るものを使用し、「感心」のアニメーションには文字が1文字ずつ上下に躍動するものをそれぞれ使用した。

なお、動画制御にはYouTube Javascript PlayerAPI[6]、音声分析にはpraat[7]、音声認識にはJulius[8]、字幕のアニメーション化にはTextillate.js[9]を利用する。以下に本研究のシ

システム構成図を示す。

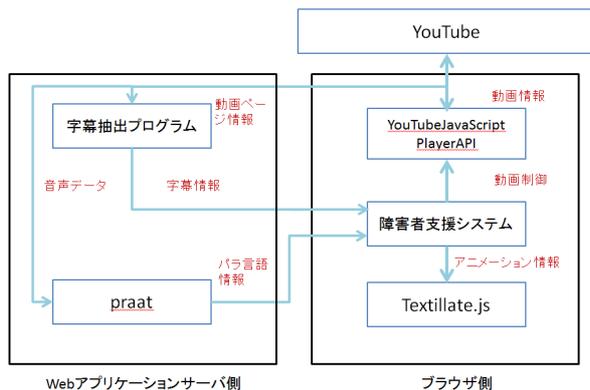


図 1: システム構成図

4. 評価

動画に「強調」, 「感心」の文字アニメーション付き字幕がそれぞれ挿入されたものを見てもらい, どのような印象を受けるかという 5 段階評価アンケートを健常者の男性 8 名に対して行った. 8 名には各設問を見せることなく動画を閲覧してもらい, 閲覧終了後に各設問に回答してもらった. 実際に聴覚障害者に利用してもらうことを想定し, 動画を閲覧する際には音声を流さず, 映像と字幕のみが流される状況とした.

以下に 5 段階評価アンケート結果の平均を求めた表を示す.

評価項目	平均
アニメーションの表示速度は妥当だと感じましたか?	3.50
強調を表現するアニメーションは妥当だと感じましたか?	4.38
感心を表現するアニメーションは妥当だと感じましたか?	2.13
アニメーションが付加された字幕は見やすいですか?	3.25

表 1: アニメーション表現調査

今回の評価においてアニメーション表示速度は 2 秒で全アニメーションの動きが完了するように設定したが, おおむね良好な回答が得られた. また, 強調に利用したアニメーションの妥当性は高評価が得られており, しっかりとパラ言語情報の視覚化が行えていることが確認できた. 一方で, 感心に利用したアニメーションからはパラ言語情報の取得がうまくできない結果

となった. これは数多くあるパラ言語情報の中から感心という言葉があまりユーザに認識されていないことと, 感心をイメージできる動きの表現が難しかったという 2 点が挙げられる. また, アニメーションが付加された字幕の見やすさについては可もなく不可もなくという結果になった. 強調については問題ないが, 感心のアニメーションは動きが大きく見づらいといった意見もあったので, アニメーション毎の動きの大きさの調整がさらに必要であると考えられる.

5. まとめ

本研究では, 話し言葉のパラ言語情報を反映したテキスト構成手法を提案し, 実装と評価を行った. アニメーション付き字幕を動画とともに見ることで, 字幕部分からパラ言語情報を取得でき, より動画内容の理解が可能となる. 今後の課題として今回の実験では聴覚障害者にシステムを利用してもらっていないので, 実際にご利用してもらう必要がある. また, アニメーションの種類を増やすことでより字幕から取得できる情報量を増加させたい.

6. 参考文献

- [1] 太田創, 竹内義則, 松本哲也, 工藤博章, 大西昇著「指示語に対応する指示対象抽出による遠隔パソコン要約筆記者支援の提案」電子情報通信学会技術研究報告. WIT, 福祉情報工学 109(467), 43-48, 2010-03-05
- [2] 日本聴覚障害学生高等教育支援ネットワーク PEPNet-Japan <http://www.pepnet-j.com/>
- [3] 有賀亮, 岸俊行, 菊池英明, 野嶋栄一郎著「授業における教師の発話の音声分析によるパラ言語情報の抽出」日本教育工学会論文誌 32(1), 13-21, 2008-06-20
- [4] 水口充, 田中克己著「文字アニメーション自動合成の試み」情報処理学会研究報告. HI, ヒューマンインタフェース研究会報告 2005(114), 97-104, 2005-11-17
- [5] Youtube <http://www.youtube.com/>
- [6] YouTube JavaScript API リファレンス https://developers.google.com/youtube/js_api_reference?hl=ja#IssuingCalls
- [7] praat <http://www.fon.hum.uva.nl/praat/>
- [8] 大語彙連続音声認識エンジン Julius <http://julius.sourceforge.jp/>
- [9] textillate.js <http://jschr.github.io/textillate/>