

THE IEICE TRANSACTIONS ON FUNDAMENTALS OF ELECTRONICS, COMMUNICATIONS AND COMPUTER SCIENCES (JAPANESE EDITION)

IEICE 電子情報通信学会 **A** 論文誌

基礎・境界

VOL. J98-A NO. 1
JANUARY 2015

本PDFの扱いは、電子情報通信学会著作権規定に従うこと。
なお、本PDFは研究教育目的（非営利）に限り、著者が第三者に直接配布することができる。著者以外からの配布は禁じられている。

基礎・境界ソサイエティ

一般社団法人 **電子情報通信学会**

THE ENGINEERING SCIENCES SOCIETY

THE INSTITUTE OF ELECTRONICS, INFORMATION AND COMMUNICATION ENGINEERS

吹き出し枠形状を利用した音声表現の字幕提示方法

紺家 裕子^{†a)} 椎尾 一郎[†]

Caption Presentation Method with Speech Expression Utilizing Speech Bubble Shapes for Video Content

Yuko KONYA^{†a)} and Itiro SHIO[†]

あらまし 近年、ビデオ映像による情報の提供が増加しているが場所などの制約により音情報が不足する場合がある。そのような映像へ向けた情報提供として字幕表示が有効であるが、既存の字幕は文字のみであり、エンターテインメントとしての音声表現は表示されない。視聴者が映像を楽しむとき、発話の文言だけでなく、音量や話すスピードなど様々な情報を総合して楽しんだり、理解を深めるが、既存の字幕ではそれらが欠如している。本研究では、フォントサイズで音量を表し、吹き出しの形状にて発話のスピードを表すことによって、音情報が不足したビデオ映像でも、雰囲気を感じ、楽しめるような吹き出し型字幕提示方法を提案する。また、システムを実装し、評価アンケートを実施した結果を述べる。

キーワード 音声表現, 吹き出し形状, 字幕表示, ビデオ映像

1. ま え が き

日本語音声に対する日本語字幕というと聴覚障害者向けの情報保障としての字幕がある。加えて近年、ビデオ映像による情報の提供が増加し、多くの場所で映像が視聴されており、音を出すことが好ましくない美術館などの静かな会場や、周りの音が大きく映像の音が聞き取りにくい電車内などでも映像が利用されており、そのような場所で利用されるビデオ映像では音がなく、字幕を付与した映像を用いることがある。しかし、既存の字幕は文字情報のみであり、映像の端、若しくは外側に表示されるため、話者と字幕が離れており、誰のせりふかわからない、映像を見ると字幕を見逃し、字幕を見ると映像を見逃すなどという課題に加え、話者の音声表現、例えば声の大きさ、速さなどを読み取るのが難しいという課題があった。視聴者が映像や演劇を楽しむとき、発話の文言だけでなく、音量や話すスピードなど様々な情報を総合して楽しみ、理解を深める。演者も音声表現を用い、感情を表現して

いる。後者の課題では、実際の話者が目の前にいる演劇などと比較すると、ビデオ映像では空気の振動が伝わらない分、音声表現による雰囲気を感じづらい。このため、映像内の雰囲気を感じ、内容を理解するために、話者の音声表現を字幕で表現することが望まれていた。

本研究では、話者の特定と見逃しの課題を演劇に適用した吹き出し型字幕[1]をビデオ映像に活用することで解決し、更に、吹き出しの形状にて話者の音声表現を付加することによって、音情報が不足したビデオ映像も雰囲気を感じ、楽しめるような字幕提示方法を提案する。また、システムとして実装し、実際のビデオ映像に提案手法を適用し、その効果を検証する。

本論文の構成は以下のとおりである。2.では提案方式のコンセプトについて、続く3.では実装したシステムについて述べる。4.では、実装したシステムを用いた評価実験と結果、考察を、5.では関連研究と従来技術についてを、6.でまとめと今後の展開についてそれぞれ述べる。

2. 提 案 手 法

本研究では、バリアフリー演劇で提案した[1]吹き出し型字幕をビデオ映像に適用し、話者の音声表現を吹き出し枠の形状で表現する提示方法を提案する。

[†] お茶の水女子大学大学院人間文化創成科学研究科, 東京都
Graduate School of Humanities and Science, Ochanomizu
University, 2-1-1 Ootsuka, Bunkyo-ku, Tokyo, 112-8610
Japan

a) E-mail: yuko.konya@is.ocha.ac.jp

2.1 吹き出し型字幕の映像への適用

吹き出し型字幕を用いることで、話者の顔の近くに字幕情報を表示することが可能となる。それによって、話者と字幕が離れており、誰のせりふかわからない、演技を見ると字幕を見逃し、字幕を見ると演技を見逃すという課題を解決できる。しかし、演劇と映像では視聴者の視点の変化を考慮する必要がある。演劇では同じ座席に座り、一定の舞台を見るスタイルであり、視線の方向や視点は視聴者が自分で決めることができる。一方映像ではズームやパンなどのカメラワーク、映像のトリミングなどの編集が行われるため、視聴者の意図にかかわらず視点が移動してしまう。例えば、発話中の人が瞬間移動したかのように動いたり、映像の外に出てしまったという状況が発生する。演劇の場合はある程度固定した場所への字幕表示で賄うことができたが、カメラワーク等を含んだ映像では吹き出しの表示方法にも工夫が必要となる。そこで我々は、カメラワークに対応するために吹き出しを人物に合わせて動かすことや、吹き出しの口先の出し方を左右の下方のみではなく上方へ向けるなど複数定義することにより話者の移動に対応することとした。

2.2 音声表現の字幕表示

音情報が不足し、音声表現による雰囲気が伝わらない課題の解決として、吹き出し枠の形状とフォントサイズを用いた表現を提案する。まんがでは、吹き出しの中にせりふが表記され、話し方の表現については、「吹き出し（風船）の工夫」によって感情表現を行っており[2]、感情を吹き出し枠の形状やフォントなどで表現している。感情表現を直接吹き出し枠の形状へ反映させることも検討したが、実際に映像を視聴する場合に視聴者が自分で演者の表現を感じることで楽しむ場合も多い。例えば、同じ「怒り」という感情でも声を荒げない淡々と話す怒りもあるし、怒鳴り散らす怒りもある。その表現も含めて視聴者が感じるものである。感情には喜怒哀楽をはじめ興奮、落胆、希望など多くの分類があり、これらは音声表現だけで表されるものではないため、音声表現のみが欠如している状況で詳細な感情を付与してしまうと情報過多となり、押し付けの演出を伝える可能性もある。もちろん、演出として吹き出し枠の形状を規定する方法もあるが、演出者や脚本者、話者など製作者の協力が必要となり、映像のみから得られる情報では作成が難しい。将来的に多くの映像に提案手法の字幕を付与する際にシステム化が難しく手間がかかってしまうという課題がある。映



図1 本提案手法による吹き出し表示の一例

Fig.1 Illustrative example of our proposal caption presentation.

像情報から、字幕情報を作成可能であれば、古い映像作品へあとから付与することも可能となり適用範囲も広がるというメリットもある。

そのため、本研究では、映像から客観的に取得できる情報から吹き出し枠の形状とフォントサイズをマッピングする方法を採用する。声のボリュームに応じてフォントサイズを変化させ、話し方の速さに応じて吹き出し枠の形状を変化させることとした。フォントタイプを変化させる方法もあるが、読みやすさを損なう可能性とテレビなど映像表示装置側のフォント対応に依存するためフォントタイプは固定とした。図1に本提案手法による表示の一例を示す。

2.3 音声表現と吹き出し枠の形状のマッピング

まず、基準状態となる吹き出し枠の形状を決めるために丸型（図2左）と角型（図2右）の吹き出し枠の形状について印象の違いを確認した。インタビュー形式で二人が会話している約1分の映像に吹き出し枠の形状のみを変化させた字幕を表示したものを作成し見やすさと印象をヒアリングした。ヒアリング対象は20代～50代の男性8名、女性2名で、一人ずつ個別にヒアリングした。見やすさについては、見やすかったものはどれかを3択（丸型、角型、同等）で回答してもらった。結果は、3名が丸型を、4名が角型を見やすいと回答し、3名はどちらも同等と回答した。見やすさについては角型の方が若干好まれたが、差がないという回答もあった。印象については、映像との適合度合いなども含め、自由に回答してもらった。全員が丸型の方が柔らかい印象があると回答し、3名からは丸型の方が会話している雰囲気が伝わるとの意見があった。2名から角型は硬い印象があり、ニュースなどの淡々と話しているところには適しているとの意見があった。



図 2 吹き出し枠の形状. 左は丸型, 右は角型
Fig. 2 Speech bubble shape type; (left) circle type, (right) square type.

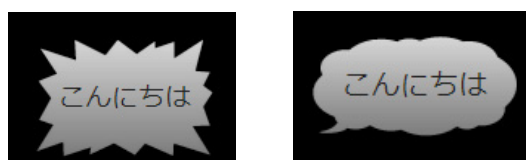


図 3 吹き出し枠の形状. 左はギザギザ型, 右は雲型
Fig. 3 Speech bubble shape type; (left) spike type, (right) cloud type.

1 名は角型が見やすいのは余白が十分に, 均等にとられているためであるという意見もあった. 本ヒアリングを受け, 会話の際には丸型を, 淡々と説明する場合には角型を用いる方法を適用することとした.

次に, 発話の速さによる変化について検討した. 激しい会話では早口で話すことが多く, 鋭角の吹き出し枠の形状でスピードを表すためにギザギザ型 (図 3 左) を選択した. まんがでも驚きや怒りなど激しい表現に用いられている. また, 逆にゆっくり話している場合は, 角の丸いフワフワした印象のある雲形 (図 3 右) を選択した. まんがでも独り言であったり, ゆったりとした雰囲気用いられている. これらの吹き出し枠の形状の効果について, 評価実験において効果を測定することとする.

3. 実 装

前記提案手法に基づき, 吹き出し枠の形状とフォントサイズを変化させる吹き出し型字幕表示システムを試作した.

3.1 吹き出し枠の形状とフォントサイズ

吹き出し枠の形状は発話の速さを表現し, 速度の速い方から, ギザギザ型, 丸型及び角型, 雲形とした. フォントサイズは発話の音量を大中小の 3 段階で規定し, それぞれ 64px, 48px, 32px に設定した. 基準とした 32 インチのテレビ画面でフルハイビジョン画像を表示し, 「中」にあたるサイズを 32px から 8px 刻みで変化させたものを 40 代, 50 代の男性 3 名に確認してもらい見づらくないサイズを決定した. 3 名とも



図 4 吹き出しの移動の一例. 右側から左なめ上へ話者が移動するのに従い, カメラが左方向へ移動している. 話者に従うようにして吹き出しも追従する

Fig. 4 An example of the movement of the balloon. Speaker in accordance to move to the upper left from the right side, the camera is moving to the left. To follow also balloon so as to follow the speaker.

加齢により視力低下を感じており必要に応じて老眼鏡を使用している. ミドル層, シニア層を含め, 比較的高年齢の人でも見やすいサイズにすることで, 幅広い年代に受け入れられやすくする工夫である. 決定した「中」のフォントサイズを基準にし, 「大」, 「小」のフォントサイズも変化させ, 見づらくなく, 差が分かる範囲として決定した.

メタデータは手作業にて作成した. メタデータに記載する吹き出し枠の形状やフォントサイズの選定は発話の速さや音量をメタデータ作成者の主観にて判断し決定した. 将来的にはしきい値を設けシステムの算出することも検討したい.

3.2 カメラワークに合わせた吹き出しの移動

吹き出しを固定した位置に表示するのではなく, 話者の動きに合わせて吹き出しを動かした. メタデータに記載する表示開始位置から表示終了位置へ表示開始時間から終了時間までの間に動かすことができる. また, 動かす軌道を移動軌道種別としてメタデータに記載しており, jQuery Easing Plugin [3] で規定できる値を指定できるようにした. この仕組みは, 編集された映像でカメラワークにより話者の画面上の位置が移動する場合や話者が移動しながら発話する場合などに活用できる. 図 4 に吹き出しを動かした一例を示す. 話者がステージの左側に向かい, 階段を上るのに従い, カメラも左側へ動く際に吹き出しの位置も画面上の人物の動きに合わせて右斜め上に移動させた.

また, シーンの切り替わりやカメラワークで画面上にいた話者が画面の外側に出てしまった場合は, 話者がいるだろう方向に向けて吹き出しの口先を画面の外側へ表示するようにした. 図 5 に示すように, 画面上部に話者がいることも想定されるため, 吹き出しの口



図 5 吹き出しの口先を上方向に向けた例（左上の吹き出し）

Fig. 5 Example of upside speech bubble tail (marked by black circle on left upside of the screen-shot).

を上部に向けることを可能にした。

3.3 システム構成

試作システムは、パソコンとディスプレイ、パソコン内にインストールした HTML5 [4] ブラウザで動作する。HTML5 ブラウザは近年テレビやセットトップボックスなどの映像受信機にも多く搭載されている HTML ブラウザで、スマートフォンなど携帯端末にも搭載されており、映像再生視聴のツールとして注目されている。試作では Chrome ブラウザを使用した。フォントサイズ、吹き出し枠の形状などの制御は JavaScript を使用し、メタデータを読み込ませることで再生している。吹き出し画像は SVG 形式 [5] の画像で規定しており、文字数やフォントサイズに応じて自動で拡大縮小するようにした。メタデータには、表示開始終了時間、表示開始終了位置、移動軌道種別、吹き出し枠の形状、フォントサイズ、字幕テキストが含まれている。字幕テキストは画面サイズにあわせて自動改行するが、読みやすさを考慮して、メタデータ内に改行を明記することで明示的に改行することができる。

試作システムはネットワークに接続しないローカル環境で実現したが、HTML ブラウザを用いているため、HTTP 通信の可能な通信回線を用いてネットワーク越しにあるサーバからデータを受信し再生することも可能である。インターネット経由で提供し、多くの人が同時に視聴することも可能となる。あわせて JavaScript によって字幕部分のみの表示非表示も制御できるため、必要な人だけにのみ字幕を表示することも可能である。映像は 1920px*1080px のフルハイビジョン映像まで対応でき、CSS で設定することにより対象となる映像に合わせてフォントサイズなどの設定変更

表 1 評価映像一覧

Table 1 Evaluation video variation.

	映像 1	映像 2	映像 3	映像 4	映像 5
吹き出し枠形状	固定	固定	可変	可変	可変
フォントサイズ	固定	可変	固定	可変	可変
音声	無音	無音	無音	無音	有音

が可能である。

4. 評価

本章では、試作システムを用いた映像を視聴してもらい、吹き出し枠の形状とフォントサイズで音声表現が伝わるかという点について評価した。

4.1 手法

被験者は 20 代後半～50 代前半の健常者男性 4 名、女性 3 名である。試作システムで字幕表示した映像をパソコンの HDMI 端子から 32 型のテレビに出力して視聴する。まず、吹き出し枠の形状を丸型、フォントサイズ一定（サイズ中と同じフォントサイズ）の映像 1 と提案手法によりフォントサイズのみを変化させた映像 2、吹き出し枠の形状のみを変化させた映像 3、吹き出し枠の形状とフォントサイズ両方を変化させた映像 4 を準備し、無音の状態で被験者が視聴した。次に映像 5 として、映像 4 の音声を聞きながら視聴した。表 1 に用いた映像の概要をまとめる。映像素材はフルハイビジョンのインタビュー映像で、メインで二人が会話をしている映像である。映像は途中で止めたり早送りしたりせず、全編を視聴した。各映像の提示時間は 4 分 45 秒である。視聴順序が影響しないように、映像 1～4 は被験者により視聴順序を変更した。

映像 1～4 を視聴後に設問 1 として「一番映像の雰囲気伝わった映像」、設問 2 として「一番見やすかった映像」、設問 3 として「一番好みの映像」を回答し、映像 5 を視聴後に設問 4 として「吹き出し枠の形状は実際の映像にあっていたか」、設問 5 として「フォントサイズは実際の映像の音量にあっていたか」を 5 段階（とても合う、やや合う、どちらでもない、あまり合わない、とても合わない）の主観評価アンケートに回答、更に、自由記述欄にてコメントや感想を得た。

4.2 結果

表 2 にアンケートの設問 1～3 の結果を、表 3 にアンケート設問 4, 5 の結果を示す。

一番雰囲気の伝わるものとしては、ほぼ全員にあたる 6 名が、吹き出し枠の形状とフォントサイズを両方

表 2 アンケート結果 設問 1-設問 3 (n=7)
Table 2 Result of questionnaire Q1-Q3 (n=7).

	映像 1	映像 2	映像 3	映像 4
設問 1	0 人	0 人	1 人	6 人
設問 2	1 人	1 人	0 人	5 人
設問 3	1 人	2 人	0 人	4 人

表 3 アンケート結果 設問 4, 5 (n=7)
Table 3 Result of questionnaire Q4-5 (n=7).

回答	設問 4	設問 5
とても合う	0 人	1 人
やや合う	4 人	5 人
どちらでもない	2 人	1 人
あまり合わない	1 人	0 人
とても合わない	0 人	0 人

変化させた映像と回答した。一方で見やすい映像や好みの映像としては、どちらも変化させない映像やフォントサイズのみ変化させた映像を好む回答もあったが、提案手法である、吹き出し枠の形状とフォントサイズを両方変化させた映像を支持している回答が一番多く、今回の評価では、提案手法が支持されたという傾向が見られた。

次に、自由記入欄の回答結果を示す。() 内はアンケート内容の確認のため著者がヒアリングの上、補足追記した文言である。

- 音がない場合は吹き出し枠の形状やフォントサイズが変わる方が(画面に)変化があって面白いが、音が聞こえると表現が誇張されているように感じる箇所もあった。
- フォントサイズが変わるだけでもインパクトがある。
- フォントサイズが変わるより、吹き出し枠(の形状)が変わった方が雰囲気伝わりやすい。
- 音がないものは、まんがを見ているみたいで面白い。
- 吹き出しの形が変わるだけでもわかるが、フォントサイズが変わるほうが更に強調されてよい。
- 文字が大きくなり、吹き出しの形もぎざぎざなので大声を出しているかと思ったが、予想よりは小さかった。
- 枠だけ変わっているときに、実際に音量は変わらないけど驚いていて、雰囲気にあっていてと思った。
- 実際の音と比べると違和感がある所もあるが、

聞こえないときは違和感なく楽しめた。

- 雲型は実際の雰囲気になってない。(声が出ていないのかと思った)
 - 雲形はほのぼのした雰囲気が出ていてよい。
 - (フォントも吹き出し枠の形状も) 変わらない方が落ち着いて見える。
 - 人物が画面外にいるときに、吹き出しの口先が外に出ていているところがよい。掛け合いが続いている雰囲気が出ている。
- 〈改善点や追加点について〉
- 音を聞いたら、なまって話す箇所があった。方言なども表現できたら面白い。
 - 吹き出しが(話者やカメラワークにあわせて)動くと思わず(吹き出しに)注目してしまい、映像から集中がそれるときがあった。
 - 吹き出しの動きが早いと見づらいかもしれない。この映像くらいなら読める。
 - 文字色を変えたりしてもよいのではないか？

また、口頭ではあるが、被験者全員からテレビ放送や映画などで用いられている画面の下に表示される帯状の字幕より見ていて面白く、字幕と映像を往復しなくてよいので視線の動きが少なく見やすいという意見がでた。

4.3 考 察

4.3.1 吹き出し枠の形状について

日本ではまんが文化も浸透しており、学校教材などでも吹き出し表示が使われているため、多くの人に認知されており親和性が高いといえる。その中で、吹き出しの表現について先入観があり、我々の意図と異なる認識をされる場合があった。まんがでは吹き出し枠の形状のうち、口先の形状についても意味がある場合がある。例えば、口先が途切れているものは心の声を表すというものである[6]。図6に口先が途切れている吹き出し(左)と、今回利用した吹き出し(右)の図を示す。1名の被験者はこの吹き出しについて、図6の二つの吹き出しを同一とみなしており、音声を聞きながら視聴した場合に音が出ていることに違和感を示した。このような認識違いも含め吹き出し枠のもつ意味が被験者によって異なる場合が予想できるため、事前に説明をするなどの対応も必要である。

あわせて、雰囲気としてはあっているが、映像音声と比較して誇張されているように感じたという意見もあった。しかし、そのように回答している被験者も音がない状態では、何も変化させないものよりは変化さ

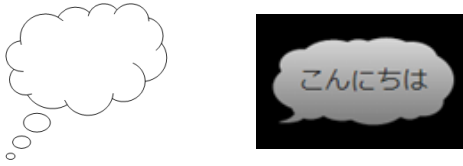


図 6 吹き出しの口が切れているタイプの吹き出し (左) と今回のシステムで使った吹き出し (右)

Fig. 6 Sample of speech bubble that has separated tail (left), and a speech bubble used in this system (right).

せた方が雰囲気が伝わっていると回答していることから微調整をしていくことで改善が可能と予想できる。また、フォントサイズのみでは違いを認識されない場合も、吹き出し枠の形状が変わると認識されるため、通常状態との違いを認知しやすいようである。

4.3.2 実際の映像との比較について

音声ありの場合と無しの場合を体験した結果、音声なしの場合では、視聴者の想像を加えて雰囲気が伝わっていたが、実際の音と聞き比べるとギャップがあることが、自由記述の回答からわかった。まんがや本など音がないものでは、個人が想像の中で雰囲気を作り出しており、その補助として吹き出し枠の形状やフォントサイズがある。想像は個人差があるため、全く同じものを再現することは難しい。個人の好みを満足させる解決策として、好みを選べるようにカスタマイズする方法や、傾向として大人数が支持する方式を採用する方法などがある。本評価では被験者数が7名と少ないため有効性については検証できないが、今回の評価結果から提案手法が映像の雰囲気を伝えることに役立つという傾向が見出された。今後、大人数に体験してもらうことで、明らかな傾向が見える可能性もあるため、更なる評価をしたい。あわせてカスタマイズ方式についても検討を進めたい。

本研究では、音声表現のうち速さと大きさに注目したが、方言などアクセントなども雰囲気を表すものである。被験者の指摘にもあったように、今回表現しきれていない箇所についても、雰囲気を伝える要素はある。例えば、音声表現の要素として声の高低や声色である。また、今回はせりふのみを表示しており、街の雑踏やBGMなどの背景音は表示しなかった。被験者から指摘はされなかったが、これらも映像の雰囲気を伝えるのに意味があると予想できるため、特徴的なものは表示することを検討したい。

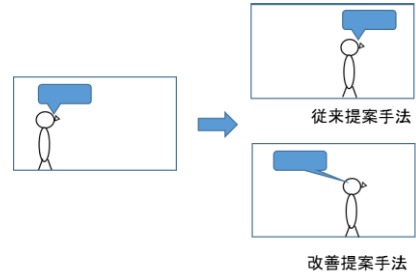


図 7 改善提案手法の模式図。話者が画面左から右へ移動するに従い吹き出しの口先のみを話者と一緒に動かす

Fig. 7 The revised proposal about a motion of a speech bubble. Only the bubble tail moves following a speaker as a speaker moves from left side to the right.

4.3.3 吹き出しの動きについて

話者が動きながら話す場合も、表情を見ながら字幕を見るほうが雰囲気が伝わりやすいと考え、話者に追従し動かす方式を採用した。今回評価に利用した4分45秒の映像中表示された116個の吹き出しのうち話者が動いている3個に適用した。早く動かすと、残像なども含め見づらくなるだろうと想定していたが、ゆっくり動かす場合でも、見づらくなるだけでなく、集中するポイントが変わってしまうというのは今回新たに得られた知見である。

改善策として、話者の近くに表示しながら、吹き出しを動かしすぎない方法として、吹き出し自体は動かさずに口先だけを話者に追従して動かす方法を提案する。改善した提案手法について図7に示す。

吹き出しの口先のみを動かす方法について、評価した結果を以下に述べる。背景は動かず人物のみが動く映像Aと人は静止しカメラがパンする映像、つまり視聴者の視点において、人物と背景が同時に動く映像Bを利用し、吹き出しを話者に追従させる場合と、吹き出しの位置は動かさずに口先を追従させる場合を比較した。動画の提示時間はそれぞれ2.5秒であり、吹き出しには13文字の文章が記載されている。6名にそれぞれの映像において吹き出し全体が動く場合と吹き出しの口先だけ動く場合を比較し、見やすい方を回答してもらった結果、映像A、映像B両者において全員が吹き出しの口先のみが動くほうが見やすいと回答した。表4に結果をまとめる。

4.3.4 システムについて

システムとしてはHTMLブラウザと通信を使うこ

表 4 アンケート結果 (n=6)
Table 4 Result of questionnaire (n=6).

	従来提案手法	改善提案手法
映像 A	0 人	6 人
映像 B	0 人	6 人

とで、同じ映像に複数種類の字幕を視聴者が選択して視聴することが可能である。個人差や好みによって選択可能となるような仕組みを提供することも可能となる。

5. 関連研究

Fels らは聴覚障害者向け字幕として放送映像を対象に感情を表すアイコンと色を用いた感情を含んだ字幕表示を提案している [7]。また、Ohene-Djan らは、子供向けの番組へ文字フォントを変化させることで感情表現をした字幕を提案している [8]。本研究とは、文字だけでは表現できない情報である感情や音声表現などを表示するという点で共通である。しかしながら、両研究とも字幕の表示は映像の下方に帯型で表示されており、話者の特定や表情と一緒に視聴することが難しい。帯型表示と吹き出し表示については、先行研究 [1], [9] や、藤井らのテレビ会議の事例 [10] などで評価されており、吹き出し表示が支持されている。また、本評価実験の被験者からも口答ではあるが、帯型より提案方式のほうが面白く、見やすいという意見を得ている。また、感情表現との対応も難しく、Fels らは演出家とともに協力して製作している。吹き出しを使った字幕表示としては、藤井らのテレビ会議等の映像に情報保障として字幕付与する研究や、人形劇への付与 [11] などがあるが、感情や音声表現を可視化しているものではない。字幕文字のフォントサイズを変えて表示する研究も行われているが [12]、読みやすさを目的とした変更であり、音量に合わせて変化するものではなく、本研究とは目的が異なっている。

吹き出しと感情表現をあわせているものとして、授業支援のシステム [13], [14] がある。吹き出しと漫符と呼ばれる線や記号を使い教員の音声表現をパソコンのディスプレイ上に表示している。話者が教員一人を想定しているため、吹き出しの口先の向きは話者を示しているものではない。生徒は手元のパソコン画面をみているため、教員の表情と合わせてみることも難しい。本研究では、画面の上に吹き出しを乗せ、話者の表情

と合わせて視聴する手法であること、話者の向きに吹き出しの口を出すことで、話者の居場所をわかりやすくし、複数話者がいる場合にも対応することができる。

吹き出しの提示場所に関する研究がある [15], [16]。これらは画像の解析や視線情報などを用いて吹き出し表示位置の最適化を行っている。本研究では、表示方法に注力したため、吹き出しの提示位置について自動化を行っていないが、今後の課題としてオーサリングの簡易化を検討するにあたり参考にしたい。

6. む す び

本研究では、音情報が不足している場合の情報を提供するための字幕として、吹き出しを用いた字幕提示方法を提案した。

提案手法では、フォントサイズで音量を示し、吹き出し枠の形状で発話の速度を表すことで、発話の音声表現を可視化し、映像内の雰囲気伝えることができる。

評価実験により、音のない状態では、単なる吹き出し表示より、提案方式によりフォントサイズと吹き出し枠の形状を両方変化した表示が映像の雰囲気を伝えることに役立つという傾向が見出された。ただし、実際の音声と比較するとギャップがある箇所があり、忠実に表現を伝えるものではなかった。動く話者に合わせて吹き出しを動かす方法においては、吹き出し自体が動くことで文字が読みづらくなったり吹き出しに注目がいくことが判ったが、吹き出しを極力動かさず口先のみで話者を追従することで改善できることが判った。

今後は、吹き出し枠の形状の工夫やサイズの微調整をすることでギャップを埋めるとともに、個人の好みに対応できるように字幕を選択可能な構成を検討し、多くの人や多くの利用シーンに対応できる仕組みを提案し、健常者だけでなく聴覚障害者、高齢者などが様々なビデオ映像をエンターテインメントとして楽しめるしくみに貢献できる手法としたい。また、実用化に向けてオーサリングの自動化などを工夫し、吹き出しの表示位置などについても最適化する方法を検討したい。

文 献

- [1] 紺家裕子, 椎尾一郎, “バリアフリー演劇における聴覚障害者向け情報表示方法,” ヒューマンインタフェース学会論文誌, vol.14, no.4, pp.9–16, Nov. 2012.
- [2] 高月義照, “まんがにおける表現技法の進化—何がまんがを文芸に成長させたのか,” 東海大学紀要, vol.20, pp.53–75, March 2011.

- [3] jQuery Easing Plugin (version 1.3), <http://gsgd.co.uk/sandbox/jquery/easing/> (2014 年 3 月 26 日確認)
- [4] W3C HTML5, <http://www.w3.org/TR/2014/CR-html5-20140204/> (2014 年 3 月 26 日確認)
- [5] W3C SCALABLE VECTOR GRAPHICS (SVG), <http://www.w3.org/Graphics/SVG/> (2014 年 3 月 26 日確認)
- [6] 小学館新コミまんが家養成講座, http://shincomi.webshogakukan.com/school/2006/09/1_4.html (2014 年 8 月 6 日確認)
- [7] D.I. Fels, D.G. Lee, C. Branje, and M. Hornburg, "Emotive captioning and access to television," Proceedings of the Eleventh Americas Conference on Information Systems, pp.2330-2337, Omaha, NE, USA, Aug. 2005.
- [8] J. Ohene-Djan, J. Wright, and K. Combie-Smith, "Emotional subtitles: A system and potential applications for deaf and hearing impaired people," Conference and Workshop on Assistive Technologies for Vision and Hearing Impairment: (CVHI), Aug. 2007.
- [9] 紺家裕子, 中谷彰宏, 佐藤 至, 椎尾一郎, "paralinguistic 表現を用いた聴覚障害者向け吹き出し型字幕提示方法," 情処学研報, vol.2013-EC-29 (4), pp.1-6, Aug. 2013.
- [10] 藤井絢子, 南条浩輝, 吉見毅彦, "会議の情報保障を目的とした吹き出し型字幕提示方式の検討 (システム)," 情処学研報, SLP-75, no.14, pp.75-82, Feb. 2009.
- [11] 楠 房子, 和田久美子, 江草遼平, 後関政史, 足立孝之, 溝口 博, 生田目美紀, 稲垣成哲, "インクルーシブデザインに基づいたインタラクティブな人形劇の開発と実践," 信学論 (D), vol.J96-D, no.1, pp.61-69, Jan. 2013.
- [12] 高尾哲康, "ユーザプロファイルに応じた字幕表示システム," 富山国際大学地域学部紀要, vol.8, pp.1-8, March 2008.
- [13] H. Nambo, S. Seto, H. Arai, K. Sugimori, Y. Shimomura, and H. Kawabe, "Visualization of non-verbal expressions in voice for hearing impaired," Computers Helping People with Special Needs, Lecture Notes in Computer Science, vol.7382, pp.492-499, 2012.
- [14] 瀬戸就一, 南保英孝, 新井 浩, 川辺 弘, 杉森公一, 下村有子, "聴覚障害学生に教員の口調と授業の雰囲気を与えるシステム," 情報処理全国大会講演論文集, pp.31-33, 2013.
- [15] B.-K. Chun, D.-S. Ryu, W.-I. Hwang, and H.-G. Cho, "An automated procedure for word balloon placement in cinema comics," Advances in Visual Computing, Lecture Notes in Computer Science, vol.4292, pp.576-585, 2006.
- [16] 澤田友哉, 豊浦正広, 茅 暁陽, "iMap を利用したフィルムコミックの自動生成," 画像電子学会論文誌, vol.42, no.3, pp.671-680, Sept. 2013.

(平成 26 年 4 月 7 日受付, 8 月 17 日再受付)



紺家 裕子 (正員)

1999 年 3 月お茶の水女子大学理学部数学科卒業。同年 4 月日本電信電話株式会社入社。2004 年より NTT サイバーソリューション研究所にて放送通信連携の研究開発に従事。2006 年 8 月よりエヌティティコムニケーションズにて IPTV とコミュニケーションに関する研究開発に従事。2010 年 4 月、お茶の水女子大学大学院人間文化創成研究科入学。博士後期課程在学中。



椎尾 一郎 (正員)

1979 年 3 月名古屋大学理学部物理学科卒業。1984 年 3 月東京工業大学大学院総合理工学研究科博士課程修了。同年 4 月、日本アイ・ビー・エム株式会社東京基礎研究所に入社。マルチメディアシステム、オフィスシステムなどのユーザインタフェースの研究に従事。1997 年 4 月玉川大学工学部助教授をへて 2002 年 4 月教授。2001 年 4 月～2002 年 3 月ジョージア工科大学客員研究員。2005 年 4 月よりお茶の水女子大学理学部情報科学科教授。実世界指向インタフェース、ユビキタスコンピューティングを中心に研究。情報処理学会、ソフトウェア学会、ヒューマンインタフェース学会、ACM 各会員。工学博士。