Vol.2011-GN-78 No.18
2011/1/22

携帯電話向けにテレビ番組の要約コンテンツを配信するためのオーサリングツールの開発

松本一則[†]　青木圭子[†]　池田和史[†]
柳原正[†]　服部元[†]　小野智弘[†]

各種の映像装置向けに TV 番組の 2 次利用が盛んになっているが，映像再生に必要な CPU 資源が十分でなく，コンテンツサイズが大きすぎるとの理由から携帯電話へコンテンツ提供は比較的困難である．そこで，筆者らは，携帯電話に向いた軽量コンテンツとして，紙芝居風コンテンツをＴＶ番組から自動生成する手法を以前開発した．今回，生成したコンテンツを基にインターネットサイトへの導線などを盛り込んだより高品質なコンテンツを作成することを目的としたオーサリングツールを作成したので，本稿で同ツールの機能を紹介する．

Development of an Authoring Tool to Create Summarized TV Program Contents for Cellular Phone

Kazunori Matsumoto,[†]　Keiko Aoki,　Kazushi Ikeda,[†]
Tadashi Yanagihara,[†]　Gen Hattori[†]　and　Chihiro One[†]

Recently, multi-use of television programs contents for various video-playing devices is becoming more popular.　However, delivery towards mobile phone devices is difficult, due to the size of the video or lack of CPU resources on the device.　To solve this problem, we have developed a method to automatically generate "picture-card show" styled contents from television programs.　In this manuscript, we introduce features of our tool which can create high-quality summarized contents by embedding guidance (links) to internet websites.

1. はじめに

各種の映像装置向けに TV 番組の 2 次利用が盛んになっているが，映像再生に必要な計算資源が十分でない，コンテンツサイズが大きすぎるとの理由から携帯電話へコンテンツ提供は比較的困難である．このため， MPEG-4 に代表されるコンテンツ圧縮技術の適用や画面解像度を下げるなどの工夫が行われることが多い．一方，時間が長くさまざまなシーンを含むコンテンツを要約することで，データサイズをコンパクトにする技術が注目されている．コンテンツ要約は多忙な利用者に膨大な映像情報の全体像を要領よく伝える技術であり，携帯電話等の利用形態を考慮すると今後ますますニーズが高くなると思われる．

一般的にコンテンツ要約は，映像をシーンチェンジや会話の切れ目で分割し，重要度の高いシーンで再構成することで得られる．ただし，スポーツ映像のように盛り上がったシーンを抽出することを重視する「ハイライト」とドラマのようにストーリーを保ちながら重要なシーンを満遍なく見つけ出してくる「ダイジェスト」とで利用する技術が大きく分かれる [1)]．また，要約対象をスポーツ映像等に限定して選手の動きを抽出することで要約品質を向上しようという試みがある一方，映像から抽出される一般的な特徴量を MPEGの符号化データから直接得ることで高速かつ汎用的な要約を行う研究 [2)]もある．しかし，いずれの場合も要約を動画で再構成する限り，要約コンテンツのデータサイズがある程度大きくなることは避けられない．

これに対し，著者らは動画中の代表画像と字幕（クローズドキャプション）のテキスト情報を組み合わせた要約コンテンツを生成する手法 [3)]を提案した．先の分類に従えば，生成する要約コンテンツは番組全編を扱うダイジェストの一種である．同手法の場合，要約生成過程におけるコンテンツ依存性が少なく，幅広く各種のテレビ番組に対応できるといった特徴がある．そして，動画を含まない要約であるので，要約コンテンツのサイズが圧倒的に小さいといった携帯電話向け配信に適した特徴を持っている．

本稿で紹介するオーサリングツールは，同手法でひとまず自動生成した要約を人手で多少編集し，最終的に品質上の問題がないことを確認するためのものである．このため，効率良く最上級の要約コンテンツを作り上げることを目標としている．以下，本稿では，まず，オーサリングツールが対象としている要約コンテンツ自体とその自動生成上の課題を説明する．次に課題を解決するために筆者らが採用した手法を紹介する．そして，オーサリングツール自体の機能紹介とその実装・評価を報告する．

[†] KDDI 研究所
KDDI R&D Laboratories, Inc.

Vol.2011-GN-78 No.18
2011/1/22

2. 要約コンテンツの概要と生成上の課題

図１は TV のニュース番組に似せて作った疑似番組から要約コンテンツを自動生成し，それを携帯電話上に表示したものである．ユーザは静止画を眺めながらその内容に沿った字幕を読み，ページを繰るようにして楽しむといった状況を想定したコンテンツとなっている．

地デジ番組は TS 形式を使って伝送されており，CAS（Conditional Access System）による復号を経て，字幕用のデータストリームから PTS（presentation time stamp）付きの字幕テキストを取り出すことができる．ただし，単にビデオストリームから取り出したフレーム画像と同時刻に表示される字幕を関連付けるようにして生成したコンテンツは高品質とは言えず、右記の点に工夫が必要である．

考慮すべき事項（課題）

1. 【不自然な字幕文章の切れ目】ニュースやスポーツ中継等のライブ番組の場合，字幕の切れ目が不自然であること場合が多い．このため．字幕の文章が読みにくくなっている．また，オリジナルの字幕文章は TV 受信機の特性にあったもので，比較的文章が短い．このため同一シーンでも字幕文章は分割表示されることが多く，オリジナルの字幕の単位でページを構成すると閲覧時に頻繁なページ送りが必要で，ユーザの操作が煩雑になる．
2. 【映像と字幕タイミングのずれ】ライブ番組の場合，字幕と映像とが同期していないため，字幕を表示している実際の時刻の画像と該当字幕を関連付けると違和感のあるコンテンツになってしまう．

(a) 番組の冒頭
字幕：こんにちは。VODニュース、キャスターの渡辺憲吉です

(b) 花火大会の話題
字幕：では今日最後の話題です。今年で8回目となりました東京湾大華火祭が8月12日に行われました

(c) 花火大会の話題(続き)
字幕：打ち上げられた花火は伝統的なきくか型花火から、コンピューターを駆使した新型花火まで、

(d) 花火大会の話題(続き)
字幕：バラエティとエンターテインメント性にあふれ、湾岸に押し寄せた約48万人の見物客を圧倒しました

図1 ニュース番組(擬似コンテンツ)の閲覧状況

Vol.2011-GN-78 No.18
2011/1/22

3. 高品質な要約を自動生成するための実装

2 章で上げた課題に対しては，オーサリングツールにおいても文献 [3]に述べた手法を基本的に採用している．ここでは，採用した手法の概要と効果について簡単に述べる．

3.1 字幕文章の整形

図 2 は生放送のニュース番組で実際にあった字幕のデータである．字幕③と字幕④のつながりや，字幕⑫と字幕⑬のつながりでは，「さらに」とか「これまで」といった単語が切れており，明らかに読みづらい．また，字幕⑥と字幕⑦のつながりでは，修飾対象「ドル」と修飾語「低い」が別れて表示されるため，一度に 1 行分の字幕しか表示されない場合，かなり読みにくいのがわかる． この読み読みにくさは文章中の文節や句を無視したことが原因とみなしてよい．そこで，オリジナル字幕文章を規定文字数内に収まるように連結していき，連結後の文章の切れ目が文節境界になるように調整する手法を採用している．

① ＞＞アメリカの中央銀行にあたるFRB・連邦準備制度理事会は
② 16日、金融政策を
③ 決める公開市場委員会を開き、これまで1%だった政策金利をさら
④ に引き下げて0から0．25%の間に目標を置き、事実上の
⑤ ゼロ金利政策に踏み切ることを決めました。
⑥ これを受けて金利が低い
⑦ ドルを売る動きが広がり、
⑧ 円相場は1ドル88円台まで
⑨ 上昇するなど、円高が進んでいます。
⑩ FRBは16日、金融政策を決める
⑪ 2日目の公開市場委員会を開き、銀行どうしが
⑫ 当面の資金をやり取りする際の金利水準を決める政策金利をこれま
⑬ での1%からさらに引き下げ、0から
⑭ 0．25%の間に目標を置くことを
⑮ 全会一致で決めました。アメリカの政策金利が
⑯ 0%台になるのは史上初めてで、FRBは事実上の
⑰ ゼロ金利政策に踏み切ることを決めました。

図2 字幕の例（オリジナル、17ページ分の字幕文章）

文節境界の検出と検出した境界の強弱を決定する方法については，形態素処理結果にパターンマッチングのルールを適用する既存手法 [4]を参考に，マッチング部分の高速化に実装上の手間をかけた．その理由は高速なマッチングが可能になると，文節判定に効果が見込めるより多くのルールを記述する余裕が生まれるからである．

さらに形態素処理の適用に際しては、表記の揺らぎ（長音とマイナス記号の置換、促音の挿入・削除・変形など）の正規化や，辞書ベースの文字誤り訂正を行う自前の前処理を実装している．

表 1 は字幕整形をした時と，しなかった時のコンテンツのページ数と閲覧に要する時間の比較である．比較に使用したのは、実際に放送された 15 分間のニュース番組である．字幕整形を行うことで字幕の文字列長は長くなり，1 ページ分を読むのにかかる時間は当然増加する．しかし，たとえ短い字幕を見た時でさえ理解するのにある程度の時間がかかる．このため，字幕整形は可読性向上に貢献しているといえる。

① ＞＞アメリカの中央銀行にあたるFRB・連邦準備制度理事会は16日、金融政策を決める公開市場委員会を開き、
② これまで1%だった政策金利をさらに引き下げて0から0．25%の間に目標を置き、
③ 事実上のゼロ金利政策に踏み切ることを決めました。これを受けて金利が低いドルを売る動きが広がり、
④ 円相場は1ドル88円台まで上昇するなど、円高が進んでいます。FRBは16日、金融政策を決める2日目の公開市場委員会を開き、
⑤ 銀行どうしが当面の資金をやり取りする際の金利水準を決める政策金利をこれまでの1%からさらに引き下げ、
⑥ 0から0．25%の間に目標を置くことを全会一致で決めました。アメリカの政策金利が0%台になるのは史上初めてで、
⑦ FRBは事実上のゼロ金利政策に踏み切ることを決めました。

図3 整形済み字幕の例（7ページで構成）

表 1 **ニュース番組に対する字幕整形の効果**

	コンテンツのページ数	閲覧に要する時間
字幕整形あり	89	平均 9.1 分
字幕整形なし	217	平均 12.0 分

3.2 音声認識を用いた字幕表示タイミングの補正

実際の字幕放送（ニュース番組）に対し字幕を表示すべきタイミングを人手で付与し，実際の表示タイミングとのずれを度数分布で示したものを図 4 に示す．同図から実際の字幕は 10～25 秒ほどの遅れて表示されることがわかる．先に述べたように、生放送の番組の場合, 字幕に付与されている PTS と映像の PTS の整合性はとれていない．

この問題に対し，オーサリングツールでは，音声認識結果の音素列と字幕から生成した読み（音素列）とのマッチングにより，生成字幕表示と映像のタイミングを補正している．図 5 は同手法のマッチング例である．「7 時のニュースです」が映像に遅れて字幕ストリームから得られたとき，字幕を音素に変換した「shichjino …」が，音声ストリームから得た周辺時刻の音声認識結果「konbanwa hichijino nyusudesu」と照合され，タイミングのずれが補正される．

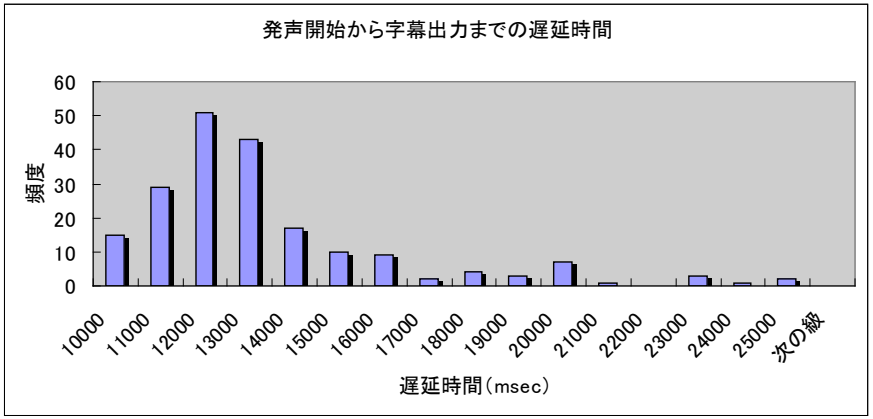

図４　実際の放送での字幕表示のずれ

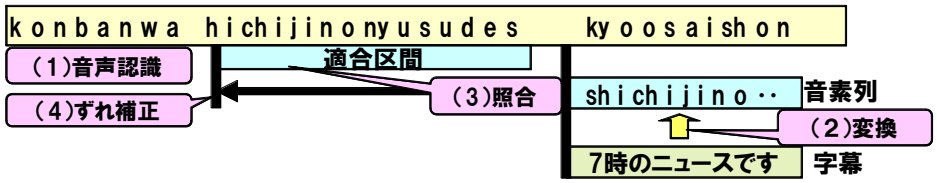

図５　字幕表示タイミング補正の例

一般的に音声タイプライタとかディクテーションとよばれるようなアプリに対しては，音声認識の能力は十分ではない．しかし，本件の場合，音声認識に成功している部分の音韻列と字幕の該当箇所がマッチングすることがある程度成功する一方，認識誤りで得られる音韻列が字幕の非該当箇所にマッチすることがあまりない. 結果的に，字幕表示タイミングを提案手法で補正した場合，マッチング誤差の平均は 38msec 程度である [3]．提案手法は十分に効果をあげている．

4. オーサリングツールの概要

開発したオーサリングツールの入力は字幕付き番組の ARIB-25 で規定される地デジの TS ストリームであり，人手で最終確認が行われた要約コンテンツを出力する．地デジのチューナーおよび B-CAS カードリーダを含めた運用も可能であり, 地デジのアンテナ等に接続すれば，RF 信号のレベルを入力とすることもできる．ただし，不特定多数の利用者を想定したものではなく，放送局等のコンテンツホルダ等による利用を想定している．ここで扱っている要約コンテンツは，TV 番組のコンテンツを改編したものになるため，要約コンテンツの配布には権利関係の調整が必要だからである．

4.1 人手による編集作業の必要性

最初に述べたように，本ツールの処理は大きく分けて，(1)地デジ番組から字幕主体の要約データを自動生成する機能と，(2)生成した要約データに対する修正と最終的な品質確認の機能に分かれる．前半の機能は 3 章で述べたような実装になっており，後段の人手による修正作業を可能な限り生じないようにことを目標に構築されている．このため，単に番組内容を理解するだけでよいとか，早見視聴が目的であれば，2 章で述べた手法で十分なコンテンツ品質が確保できる．しかし，最終的に配布されるコンテンツに対し，どうしても編集作業が必要になる場合がある．例えば，コンテンツの性格上，番組全編に対する権利調整が大変であるため，権利調整が困難な映像部分が要約コンテンツに含まれないように要約内容の編集を希望することもある．また，権利関係上の問題以外でも，コンテンツホルダが最高品質の要約を提供することを希望すれば，要約コンテンツで使用される静止画の入れ替えや，字幕文章を当初と異なるものに置き換える必要が生じる場合がある．

こうした編集要望に対して，効率的な編集と確認ができるようにオーサリングツールを提供することは，要約コンテンツの流通に寄与する．

4.2 機能の紹介

図 6 は，オーサリングツールを使用で想定している作業の流れである．利用者はとツールとのインタラクションは主として同図の「③編集・プレビュー」の段階で行わ

Vol.2011-GN-78 No.18
2011/1/22

れる．この段階で，単なる要約コンテンツの編集以外に，要約コンテンツの各ページからインターネットサイトへ導線となる web リンクの情報（ジャンプ先の URL 等）が設定も可能になっている．

図 7 は，編集中の画面の例である．周辺フレームも含めた画像と字幕の対応を見やすく表示している．タイムラインと呼ぶフレーム画像をライン上に等間隔で表示したもの内，要約コンテンツとして選択された画像については，その字幕を右側に表示している．字幕を選択すると、対応する選択画像を拡大表示したり，テキストの内容を修正することが可能になっている．また，字幕を任意の位置に挿入することも容易になっている．

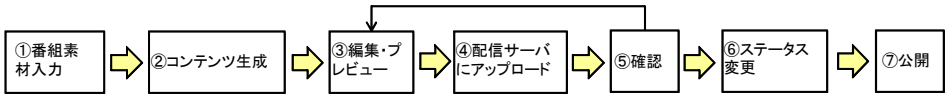

図 6　オーサリングツール利用時のワークフロー

図 7　編集画面

Vol.2011-GN-78 No.18
2011/1/22

5. おわりに

本稿では，地デジ番組を基に携帯電話向けコンテンツに適した要約コンテンツを作成するオーサリングツールを紹介した．要約コンテンツは，動画中の代表画像と字幕情報を組み合わせたものであり，コンテンツのデータサイズも小さくてすみ，手軽にTV 番組の内容が理解できるものである．オーサリングツールは，単に人手だけでコンテンツを作成するのではなく，字幕テキストの整形や音声認識による字幕と映像とのマッチングによって，入力された素材からある程度高品質な要約コンテンツから編集始めることができるようになっている．

一方，ツールには，周辺フレームも含めた画像と字幕の対応表示機能等，より高品質な要約コンテンツを配信することを目的とした編集機能も実現されている．ツールで作成した要約コンテンツを展示会などで紹介したところ，コンテンツ品質に対しては好意的な意見が得られた．

謝辞 本ツールの実現および, CEATEC 等での実証デモに協力いただいた関係各位，特に FMBC 推進グループの沖本彰に感謝します．また，日ごろから研究活動に日頃ご指導いただく KDDI 研究所伊藤泰彦会長，秋葉重幸所長，中島康之副所長，および菅谷史昭執行役員，滝嶋康弘執行役員に深謝いたします．

考文献

1) 滝嶋, "知っておきたいキーワード 映像の自動要約技術", 映像情報メディア学会誌 Vol. 62, No. 5, pp. 714-716, 2008

2) M. Suganao, Y. Nakajima and H. Yanagihara, "MPEG Content Summarization Based on Compression Domain Feature Analysis", IT. Com. 2003., SPIE 5242, 32, pp.280-288 (Sept. 2003)

3) 松本，内藤，帆足，呉，滝嶋, "地デジ放送からの字幕主体の携帯電話向け要約コンテンツの自動生成", 電子情報通信学会画像工学研究会, vol. 108, no. 425, IE2008-214, pp. 59-63, 2009年2月

4) 高梨克也・丸山岳彦・内元清貴・井佐原均. 「話し言葉の文境界 --CSJ コーパスにおける文境界の定義と半自動認定--」.『言語処理学会 第 9 回年次大会 発表論文集』, 521-524. 言語処理学会(2003).