オンライン動画コンテンツに対する ボイスコメント合成再生システムの開発

田村和也[†1]　川合康央[†1]

概要：本研究は，あらかじめ録画されたオンラインによる音楽ライブなどの映像コンテンツに，視聴者の掛け声や歓声，コメントなどのボイスコメントを追加し，複合した音声コンテンツを作成・再生するシステムを開発したものである．試作したシステムでは，音声コメントの重なりやノイズ，合成される音域の大きさなどにより，聞き取りにくいケースがあることが明らかとなった．そこで本稿では，ボイスコメントを加工処理することによって，音声情報の明瞭度を向上させることができるかについて検討を行うこととする．

1. はじめに

現在，動画画投稿サービスは，多くの人々に親しまれているコンテンツであり，今後も需要が見込める Web サービスである．また，2019 年以降，世界的な課題となっている新型コロナ感染症に対し，これまでライブハウスなどで行われてきた音楽ライブなどの開演が困難となりつつある一方で，オンラインでの動画配信などが着目されている[1]．動画投稿サービスの例としては，YouTube[2]や TikTok[3]等のコンテンツが挙げられる．また，同じく動画投稿サイトであるニコニコ動画[4]では，画面上の動画に視聴者のコメントがテキストによって重畳表示されるといった，独自の特徴的なインタフェースを提案しており，これは海外の動画サイト BiriBiri [5]などでも採用されている．

本研究では，これまでにユーザコメントを文字ではなく音声として投稿できる新たな動画投稿サービスの検討と，その可能性について，試作システムの開発を通じて検証を行ってきた．本稿は，試作システムの評価の際に課題として挙げられた音質の改善についての報告を行うものである．

動画視聴時に，ユーザの音声を扱うことについて，松長ら[6]は，WakWakTube のユーザアバターに音声情報を加える実験を行っており，それが効果的に作用していることを示している．また，高野ら[7]は，他ユーザと共に TV を視聴しているような臨場感を持つ動画視聴システムを提案しており，会話が成り立つ精度での再生同期制御を実現した．本研究では，これら先行研究を踏まえ，独自のボイスコメントの合成再生システムの開発を行う．

2. ボイスコメント合成再生システムの開発

2.1 開発環境

今回作成したシステムの開発では，開発言語として，HTML，PHP，JavaScript を用いた．HTML で，Web サイトのインタフェースを，PHP でサーバー側の処理を，JavaScript でクライアント側の処理を行った．開発環境は VisualStudioCode，XAMPP（Apache），chrome を用いた．開発したシステムの画面を示す（図 1）．

図 1　開発システムの動作画面

2.2 動画に対する音声コメントの取得

視聴者による音声コメントは，クライアントが PC に設定しているマイクデバイスから取得するものとした．入力された音声は，Web ページに埋め込まれた JavaScript によ

1 文教大学

って処理され，音声ファイルとして出力される．この時，音声ファイルには，動画のどのタイミングでコメントされたかという時間情報と，ユーザを判別するための識別番号が記録されているものとした．クライアントは，これらの情報を投稿することができ，その音声ファイルは PHP によってサーバー上に記録される（図 2）．

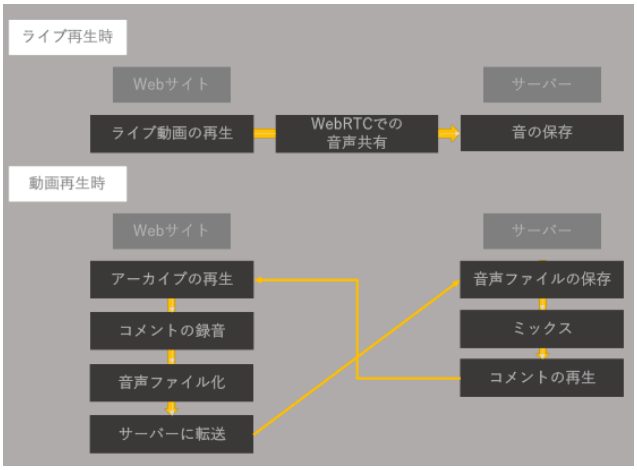

図 2　システム構成図

2.3　動画に対する音声コメントの再生

動画再生中は，常に現在の再生時間を JavaScript にて取得する．そして，その時間に対応した音声コメントが投稿されている場合，保存されたコメントを重ねて再生する．コメントの音量は，最終的にはクライアントが調節できるものとしたが，デフォルトでは元の動画の音量の半分ほどに設定した（図 2）．

2.4　ライブ配信に対する音声コメントの取得

音楽イベントなどのライブ配信の場合，リアルタイムに音声が伝わらず臨場感を味わえないと考えられる．そのため，本システムでは WebRTC を使って，グループ通話に近い形で，ライブ配信機能を用意することとした．今回は，Skyway という API を利用して実装した．

3.　実験と考察

3.1　音声処理の改善

試作したボイスコメントの合成再生システムは，ボイスコメントを重ね続けていくと，元の音声情報が薄くなり，やがて雑音と化していった． この問題は，ボイスコメント内の雑音，合成された音域の大小によって，聞き取りが困難になっているのだと考え，音声情報を修復するいくつかの手法を試行した．結果，手動でノイズとなる要素を取り除いた際に，音質に改善されたため，これらが原因の一部だと確認された．

まず，雑音について，見ていくこととする．雑音については，一定程度以下の大きさの音を取り除くことで改善される．だが，この雑音除去処理を自動で行う場合，コメントごとに雑音を検出しなければならないため，その判別は非常に難しい．AI を用いた判別処理も検討したが，雑音の種類は，録音された環境に応じて多岐にわたるため，本システムでは単純な工程を採用することとした．元となる動画の音声とは異なり，ボイスコメント自体に音質はそこまで求められないと考え，本システムでは，暫定的にすべての音量の 1/4 の大きさをカットし，3/4 を引き延ばして雑音の低減を目指した．この 1/4 という比率は，いくつかの音源をもとに手動で調整した結果による数値であるが，今後，被験者実験を行い，最適な値を明らかにしていくこととする．

次に，ボイスコメントの音量を調整して合成を行った．この処理によって，コメント間の音量の差異が軽減され，聞き取りやすさが向上した．ただし，元となるボイスコメントの音量が小さい場合，引き延ばした際に音が歪んでしまうことがあるため，今後対応策を検討したい．

最後に音域の問題がある．動画の音域とコメントの音域が重なると，聞き取りが困難になることが分かった．この問題に対応するため，すべてのコメントの音域を，元の動画の音域よりも，音程を高くする方法を採用した．この処理によって，コメント音声は人工音声感が強くなるが，実際の肉声よりも匿名性が高くなるといった点では，本システムの利用シーンを考えると適切である可能性がある．さらに，各コメントの音程を適度に散りばめれば，コメントごとの聞き分けも可能となる．今後も，引き続き評価を行っていくこととする．

図 3 は，加工前（上）と加工後（下）の音声波形を比較したものである．比べてみると，加工前より幅が狭くなだらかな波形になっていることが確認できる．加工された音質は，加工前のものと比較して，聞き取りやすいものとなった．

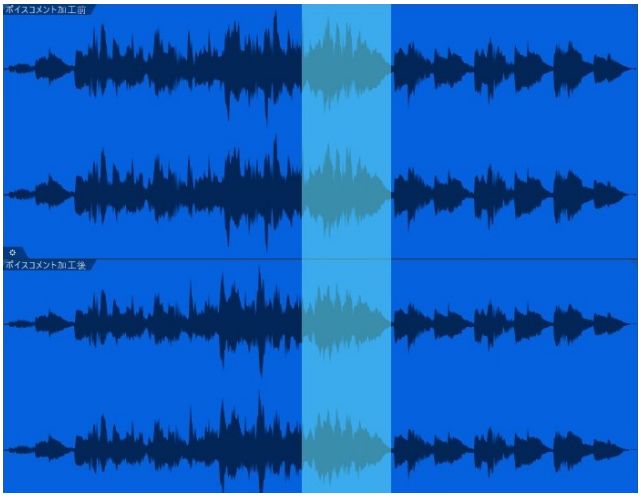

図 3　波形の比較：加工前（上）と加工後（下）

4. まとめ

本研究は，動画につけられた音声コメントを合成して再生するボイスコメント合成再生システムの開発を行ったものである．試作システムから，いくつかの課題が明らかとなったため，システムの改修を試み，音質向上などの改善を行うことができた．一方で，ノイズの閾値など，今後評価実験を含めて改良を行い，システムの精度を向上させる必要がある．また，より適切な音声データの混合方法も工夫する必要がある．これらの課題に対して，今後も被験者を用いた評価実験を繰り返し，改善していきたいと考えている．

謝辞 本研究は JSPS 科研費 JP 19K12665 及び JP20K12517 の助成を受けたものです．

参考文献

[1] 三浦文夫. オンライン配信ライブコンサートに関する課題の整理. 関西大学社会学部紀要. vol.53, no.1, 2021, p.185-201.
[2] “YouTube”. https://www.youtube.com/, (参照 2021-12-22).
[3] “TikTok - 見つけよう、 次の瞬感を。”. https://www.tiktok.com/ja-JP/, (参照 2021-12-22).
[4] “ニコニコ”. https://www.nicovideo.jp/, (参照 2021-12-22).
[5] “哔哩哔哩 (゜-゜)つロ 干杯~-bilibili”. https://www.bilibili.com/, (参照 2021-12-22).
[6] 松長雄也, 谷中俊介, 坂内祐一. 動画視聴における他視聴者の音声情報の再生方式,マルチメディア, 分散協調とモバイルシンポジウム 2018 論文集, 2018, p.530-533.
[7] 高野祐太郎, 大島浩太, 田島孝治, 高田治, 寺田松昭. 投稿型動画視聴におけるユーザ間リアルタイムコミュニケーション支援システム,電子情報通信学会論文誌 D, vol.93, no.10, 2010, p.2302-2316.