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　２８　年　　６　月　３０　日現在

機関番号：　５２３０１
研究種目：　若手研究(B)
研究期間：　2012～2015
課題番号：　２４７００１９１

研究課題名（和文）発話アニメーションにおけるリップシンクの逐次出力に関する研究

研究課題名（英文）A study on an incremental lip-sync technique for speech animation

研究代表者
川本　真一（KAWAMOTO, SHINICHI）

群馬工業高等専門学校・電子情報工学科・講師

研究者番号：７０４１８５０７

交付決定額（研究期間全体）：（直接経費）　　3,400,000円

研究成果の概要（和文）：出力に遅延が許容される利用環境を想定し、入力音声を一定時間遅延させて出力した音声と同期するアニメーション出力を対象に、音声入力と並行して、入力音声に対する視覚素認識結果の漸次的な出力から、視覚素ごとに設計されたフィルタを利用して口形状の混合重み系列を出力し，ブレンドシェープ法（形状の線形和モデル）によりリップシンクアニメーション（音声と同期した唇の動き）の逐次出力を実現した。

研究成果の概要（英文）：dependent filtering and incremental viseme recognition. This method has a simple customization technique of mouth movement in consideration of mouth movement velocity without a multi-modal database between speech and mouth movement for training. In our approach, a speech signal and a CG character data were given as inputs. This system outputs blending weights of each mouth shape based on blendshapes, which is basic technique of animation and widely used in CG software. First, we convert speech to a viseme sequence on the fly using a viseme recognizer. Then, we apply viseme-dependent filters for generating blending weights. Finally, Lip-sync animation is generated using blendshapes with calculated blending weights. As a result, the proposed method can synthesize incremental lip-sync animation with almost 300 ms delay, and synchronize mouth movement along with the speech with the same delay as input speech.

研究分野：　音声情報処理

キーワード：　リップシンク

１．研究開始当初の背景

音声と映像によるマルチモーダル（複数の伝達手段の組み合わせた）コミュニケーションは，人間が用いる最も基本的な情報伝達手段の一つであり，人間同士のコミュニケーションのみならず，人間と機械とのインタラクションや，機械を介した人間同士のコミュニケーションなど，人間が関わるメディアにおいて重要な役割を果たしている．ゲームやアニメ等のコンテンツ制作も，コンテンツプロバイダからユーザに対して音声・映像を介して情報を提供するという意味で，音声・映像による（単方向の）情報伝達手段の一つととらえることができる．アニメやゲーム等ではＣＧ（コンピュータグラフィクス）技術を活用したキャラクタアニメーションが幅広く利用されている．また，対話システムなど即応性 (入力に対して即座に応答できること) が重要であるインタラクティブなシステムへの応用も多い．人間と見間違うようなリアリティの高いＣＧキャラクタのアニメーションだけではなく，アニメやゲームに登場するような抽象度の高い，デフォルメされたキャラクタに対するアニメーションの需要も高く，対話システムなど人間とのインタラクションを必要とするシステムにおいて，リアリティの高いＣＧキャラクタを採用しない事例も増えてきている．このようなＣＧキャラクタを介して音声コミュニケーションを実現する，もしくはＣＧキャラクタが音声を発話しているような状況を作る際に，発話アニメーションの生成技術は自然性を確保する上で重要である．

発話アニメーションのような音声とＣＧキャラクタアニメーションの同期ずれは，アニメーション全体の自然性に大きく影響する．特にリップシンク（音声に同期した口形状アニメーション）は発話アニメーションの最も基本的な要素であり，同期ずれが顕著に表れる．また，リップシンクの同期ずれが自然性や了解度に影響することも知られている．また，抽象度の高い，デフォルメされたキャラクタに対するアニメーションにおいては，必ずしも人間から観測した動きを高精細に再現することはなく，キャラクタの見た目に合わせた動きの簡単化が可能であることが望ましい．

申請者は本研究の基盤となる，口の動きの単純化に着目した，抽象度の高いキャラクタに対するリップシンクのための要素技術を提案し，いくつかの映像制作における実証実験により，アニメーション制作現場におけるリップシンク制作支援技術としての有効性の実証を進めてきた．

このような背景のもと，本研究では，音声発話と並行して処理を行い，その結果をリップシンクとして反映させるため，アニメーションの簡単化に着目した基盤技術を拡張し，音声認識の途中結果に基づいて，リップシンクのためのキーフレームを逐次出力する方法へと拡張する．

２．研究の目的

若干の遅延を許容する利用環境を想定し，キャラクタの見た目に合わせて口の動きの簡単化をシステム設計者がカスタマイズ可能な枠組みを有し，音声入力と並行して，発話アニメーションの逐次出力技術を実現することを目標とする．

３．研究の方法

以下に示す３つの観点から研究を展開した．

（１）奥行き情報の影響

画面を介した対面対話を実現することを想定した場合に，奥行き情報を両眼の視差によって明示的に与えることが音声の聞き取りの改善に大きく寄与するかについて把握するため，立体視視聴におけるリップシンクの影響を雑音重畳音声の聞き取り実験により調査し，奥行き情報をシステムに取り入れる優先度について検討した．比較として，両眼の視差を利用せず通常の映像視聴を想定した場合と比較することで，奥行きの情報が音声の聞き取りの了解度改善に大きく寄与しうるかを調査した．また，参考として映像を提示しない音声のみの条件も実験対象とした．

（２）音声駆動による漸次的発話アニメーション出力

音声入力と並行して，漸次的な音声認識の結果から，口形状アニメーションを生成する処理を検討した．その際，動きの簡単化を実現できる枠組みを組み込むことを念頭に置いたシステム設計および実現法を検討した．また，基盤技術においてアニメーションの方式として採用していたブレンドシェープ法（形状の重み付き線形和モデル）を，今回の手法においても継承し，発話アニメーションを想定した基本口形状（視覚素）がモデルとして用意されていることを前提とした．これにより，問題の対象を「音声入力に対して漸次的に視覚素に対応する重み系列を生成する」ことに設定した．

（３）映像に対する音声の同期発声との比較

口の動きの同期の観点から，出力アニメーション，および遅延量について検討した．比較対象として，（声優などの特殊なスキルを持つ人ではなく）一般の人が音声に合わせて発話した際の，ずれの量を参考とすることで，人がリアルタイムに口パクをする際の定性的な同期精度との比較を対象とした．

４．研究成果

（１）奥行き情報の影響

臨場感を高める技術として両眼視差立体視を利用した3次元映像が映画やゲームなどのコンテンツ普及に伴い身近になりつつある．通常の映像とは異なり，3 次元映像は視差による奥行き情報の表現が可能である．3 次元映像の視差による奥行き情報が音声コミュニケーションに有益であるかを探るため，雑音重畳音声の聴取実験により，奥行き情報の提示による音声了解度への影響を検討した．音声了解度を調べるため，4 桁数字音声の聴取実験を実施した．提示音声には白色雑音を重畳し，提示映像には，音声と同時に発話者正面から撮影したものを使用した．比較のため通常の映像提示では撮影した3次元映像の左レンズの映像を両眼に見せるようにし，音声のみの提示では画面に何も表示せずに実験した．SNR 2 条件，映像 3 条件の全組み合わせに対し，4 桁の数字列 25 個(計 2×3×25=150 個)を各実験協力者に提示し，その数字正答率を比較した．
3 次元映像提示条件の正答率は，通常映像と比べて有意差は見られなかった．映像内の発話者を注視して発話内容を聞き取ろうとするとき，頭を動かし視点を変えながら聴取することは少ないと考えられ，結果として奥行き情報が有効に活用されなかったことを示唆する結果となった．

（２）音声駆動による漸次的発話アニメーション出力

発話者と聴取者の役割が明確であり，遅延を許容する片方向の通信環境下での使用を想定した音声駆動による漸次的発話アニメーション手法を対象に検討を進める．このとき発話アニメーションを想定した基本口形状（視覚素）のみが共通であるとし，発話者の特徴とアニメーション対象の顔の特徴は必ずしも一致させる必要はなく（視覚素認識系とアニメーション生成が独立している），許容する遅延時間に応じて言語的な制約を視覚素認識に取り入れることを可能としたシステムを設計し，プロトタイプシステムを実装した．視覚素認識系とアニメーション生成とを独立させることで，視覚素定義さえ一致させれば様々なアニメーション生成系を使用することが出来，リアルな顔画像のアニメーションのみならず，カートゥーンキャラクタなどにも適応が可能である．その際，音声特徴量と口形状との時間的に同期のとれた大量の学習データ（パラレルコーパス）も必要としない．さらに，言語制約により視覚素認識の精度向上が期待できる．
入力音声に対して視覚素認識を行った漸次的な認識結果に基づき，視覚素ごとに設計されたフィルタ（視覚素依存フィルタ）によって，視覚素ごとの重み系列に変換し，その重み系列からブレンドシェープ法（形状の重み付き線形和モデル）によって発話アニメーションを出力する． 視覚素依存フィルタの形状を変更することで，視覚素間のわたりでの口形状の遷移の仕方をカスタマイズすることが可能である．
プロトタイプシステムを用いた動作検証により，入力音声に対して 300 ミリ秒程度の遅延で発話アニメーションを生成できること，言語的な制約を導入することで特に子音の漸次的な視覚素認識結果が改善することを確認した．これは，短時間であっても言語的な制約を加えることで，発話アニメーションの精度向上につながることを示唆するものである．

（３）映像に対する音声の同期発声との比較

音声駆動の発話アニメーションを人が漸次的に行うような状況ではどの程度の遅延が生じるかを調査した．例えば人が音声を聞きながら口パクをするような状況に当てはめて考えるとき，仮に，発話内容が予測出来たとしても，アフレコで声を収録する状況と同じく，経験に応じた遅延が生じることとなる．一般の人が映像に対してアフレコにより音声を収録した事例における，音声の遅延についても分析したところ，発話区間レベルでは73.8%, 音素レベルでは 79.5%の発話がリップシンクにおける許容限内の遅延で収録できていることを確認した．上記（２）において提案した音声駆動による漸次的発話アニメーション手法と比較すると，人によるアフレコの方が発話内容を予測できる（つまりより強い言語制約を利用できる）という点では有利であり，視覚素認識誤りの質は異なるため厳密な比較は困難である．しかし，単純に発話のずれをフレームレベルの視覚素の誤差として置き換えて比較すると，その誤差は提案システムで 72.1%の視覚素正解率，人によるアフレコでは 79.5%の正答率（音素に対応する視覚素は１つであるとする）であり，提案システムに改善の余地が残されていることを示唆する結果といえる．今後は，システムのさらなる改善を目指し，発話内容の予測も含め，さらに検討を進める必要がある．

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　２　件）

① 川本真一，視覚素依存フィルタによる漸次的音声駆動発話アニメーション，電子情報通信学会論文誌，査読有，Vol.J97-D, No.9，2014，pp.1416-1425
② 川本真一，森島繁生，中村　哲，VoiceDub：複数タイミング情報をともな

う映像エンタテイメント向け音声同期収録支援システム, 情報処理学会論文誌, 査読有, Vol.56, No. 4, 2015, 1142-1151

〔学会発表〕（計　２　件）

① Shin'ichi Kawamoto , Speech-driven realtime lip-synch animation with viseme-dependent filters , ACM SIGGRAPH 2013 Posters , 2013.7.21-2013.7.25 , Anaheim California USA

② 川本真一, 雑音重畳音声の了解度における３次元映像提示の影響, 電子情報通信学会２０１３年総合大会, 2013.3.21, 岐阜

〔図書〕（計　０　件）

〔産業財産権〕
○出願状況（計　１　件）

名称：発話アニメーション生成装置，方法，及びプログラム
発明者：川本真一
権利者：国立大学法人北陸先端科学技術大学院大学
種類：特許
番号：特願 2014-147933 号
出願年月日：2014-07-18
国内外の別：国内

〔その他〕
http://www.gunma-ct.ac.jp/gakka/09-03-15.htm

６．研究組織
(1)研究代表者
川本　真一　（KAWAMOTO　SHINICHI）
群馬工業高等専門学校・電子情報工学科・講師
研究者番号：７０４１８５０７

(2)研究分担者
なし

(3)連携研究者
なし