様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

令和　２ 年　 ６ 月 ３０ 日現在

機関番号：３２６５７
研究種目：基盤研究(C)（一般）
研究期間：2017～2019
課題番号：１７Ｋ００２８１

研究課題名（和文）話者の状態に適応する話速変換会話システムの音声・映像処理とそのモデル化

研究課題名（英文）Speech and video processing and modeling of speech rate conversion conversation system adapting to speaker state

研究代表者
斎藤　博人（Saito, Hiroto）

東京電機大学・システムデザイン工学部・准教授

研究者番号：００３２８５１９

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）：話速変換を用いる会話において，被支援者がタイムラグを知覚しない映像と音声の処理手法を実現した．被支援者は，映像と音声とが同期するリップシンクが取れる会話インタフェースで会話が可能となった．また，ゆっくりとした発話の聴取の支援を必要としない「話し手」に対する支援機能として，話し手が，聞き手側のゆっくりとした発話の受聴内容を把握できるように，話し手の発話終了後に話し手自身にループバック音声（音声フィードバック）を適応的に再生するインタフェースの実装し，音声フィードバックの有効性を明らかにした．

研究成果の学術的意義や社会的意義
　遠隔ミーティングや遠隔授業等の利用機会が増加している現在，多人数が参加する会話場において，ゆっくりとした発話で聴取したい参加者もいることが想定される．
本研究成果は，聞き取りに支障のある弱者が他者との対等なコミュニケーションをとるためのインタフェース設計をした．これにより，これまで周囲とのコミュニケーションが遠慮がちで孤独感を持っている人でも，会話の場に加わることができ，例えば高齢者の生活の質の向上や，非母語の学習者の言語の能力向上につながる．本研究成果は，人間同士のコミュニケーションの多様な場に応用が可能である．

研究成果の概要（英文）：We realized a video and speech processing method in which the support recipient does not perceive time lag in conversation using speech rate conversion system. The hearer is able to have a conversation with a conversation interface that could take lip sync in which video and speech are synchronized.
In addition, as a support method for the "speaker" who does not need to support listening to the slow utterance, the speaker listen the loops back speech himself after the utterance so that the speaker can understand the listening contents of the slow utterance on the hearer side. We implemented an interface that adaptively plays loop-back voice (voice feedback) and clarified the effectiveness of voice feedback.

研究分野：会話インタフェース

キーワード：話速変換　順番交替　遠隔コミュニケーション　遅延

様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９（共通）

１．研究開始当初の背景

声質を変えずに音声を時間伸長する話速変換は，非母語で会話をする学習者や認知機能が衰えた高齢者の聞き取りの支援に有効な技術である．しかし，人間同士のインタラクションに話速変換を適用すると，音声の時間伸長によって生じる話し手と聞き手のタイムラグが原因となり，発話衝突やきまずい沈黙が生じる問題があった．話速変換会話における課題は，順番交替時に発生する発話衝突と気まずい沈黙の回避にある．話速変換を利用する会話では，聞き手の聴取時間が話し手の発話時間よりも長くなるため，話し手は，聞き手がいつ聞き終わるのかが分かりにくくなる．その結果，複数人会話では，次の発話が誰によっていつ開始されるのか予測するのが困難になる．また，聞き手は，話し手によって既に話された発話に，後続する発話があるのかどうかの判断が難しく，慎重に自らの発話開始を待つようになり，これがタイムラグに起因する発話衝突の要因になっていた．このタイムラグは，衛星電話のような定遅延にはならず，発話時間長に比例した時間長になるため，参与者によるタイムラグの長さの予測を困難にしている．さらに，会話では「話し手」，「(主たる) 聞き手」，「傍参与者」の役割が時々刻々と遷移するため，話速変換会話のトラブルを回避するためのインタフェース設計には「話し手」，「聞き手，傍参与者」の参与役割ごとに設計が必要になる．したがって，話速変換を用いる会話では，参与者に対してそれぞれの立場に適した現在の発話に対するフィードバック，および次の発話に対するフィードバックをシステムが提供し，参与者らの発話(発話開始，発話速度）のマネージメントが課題となっていた．

２．研究の目的

本研究では，話速変換を用いる複数人会話における弱者（ゆっくり再生の支援が必要な被支援者）と，健常者（ゆっくり再生の支援が必要が無い人）との会話において，弱者に対する協力を健常者から違和感なく自然に引き出すインタフェースを開発して，会話参与者全員が共有できる公平な会話の場を構築することを目的とする．具体的には，以下の３つを達成目標とした．

(1)ゆっくり音声を受聴する弱者（被支援者）がタイムラグを知覚しない話速変換会話システムの実現

(2) 弱者に対する協力を健常者から違和感なく自然に引き出すインタフェースの開発

(3) 上記モデルを統合した話速変換会話インタフェースの構築とその評価

３．研究の方法

「ユーザとシステム」のインタフェースの評価は工学的アプローチに基づく定量分析を採用し，モデル化したタイムラグによるトラブルの回避手法が，ユーザに対して的確かつ有効に伝達されているかのシステム評価を実施する．システム設計の途中段階においても，プロトタイプで会話実験を実施し「ユーザとユーザ」による話速変換会話の質の評価をする．

(1) ゆっくり音声を受聴する弱者（被支援者）がタイムラグを知覚しない映像と音声の支援とその制御

話速変換システムの被支援者，すなわちゆっくり音声を受聴する弱者には，音声と同期して映像も伸長し，ゆっくりした時間進行が現実世界のものと知覚できる会話インタフェースを設計する．この課題では，話速変換音声と映像処理を同期する処理のモデル化と，それによって生ずる発話終了直後の映像がコマ落ちする構造的問題の解決法を策定する．この問題は「ユーザとシステム」を評価し，評価者は独立にそのインタフェースのユーザビリティ評価をする．なお，映像がコマ落ちする区間は，話し手にとってはタイムラグ（フィードバック情報受信中）の時間に相当する．したがって，この区間の映像には話し手の意図の表出行動は含まれていないと予測される．

(2) 話速変換の会話中に生じるタイムラグを補完するユーザインタフェースの策定

話速変換における話し手へのフィードバックは，発話中と発話後（発話終了から聞き手の聴取終了の時間）の両方が可能である．発話中は妨害にならない視覚フィードバック，発話後は強制力が強い音声フィードバックが有効と考えている．しかしながら，発話中の視覚フィードバックでは，タイムラグのみを可視化する効果は限定的になることも予測できる．そこで，システムが話し手の発話のモーラ数を自動計測し，その数値に対して話速の伸長率を適応的に決定して音声を処理する．そうすればゆっくり話すようになった話し手（健常者）の発話はそのまま相手に伝わる．これを実現するためには，自分の話したモーラ数の累積と相手に伝えられたモーラ数の累積の差分を「視覚フィードバック」で話し手に提示する．

４．研究成果

(1)　話速変換を利用する会話における映像同期手法とその効果について

ゆっくり音声を受聴する被支援者がタイムラグを知覚しない映像と音声の支援とその制御として，ゆっくりとした発話に変換された音声を聴取するユーザに対して，映像も音声と同期するゆっくりとした映像に加工し，リップシンクが取れる会話インタフェースの実装とその評価を行った．映像も音声と同期して，リップシンクが取れる会話インタフェースでは，話し手が実際に

ゆっくりと話しかけていると知覚できる．その結果，話し手の発話に対して，時間差を意識せずに自然なタイミングで応答が可能となる．しかしながら，ゆっくりとした音声に同期して，映像もゆっくりとした再生にすると，利点がある一方で，累積する遅れ時間を現実時間に復帰させるために映像の一部を削除し，意図的に欠落（コマ落ち）させる必要があった．コマ落ちが生ずる区間について図１で説明する．図の発話では，話し手の発話区間はビデオフレーム「V1」から「V4」までである．したがって，「N5」以降の映像フレームは非音声区間の映像となり発話による口の動きはない．発話された区間の音声・映像を共に 1.5 倍の時間伸長処理をすると，聞き手側では，現実時間の映像フレーム「N6」で聴取終了となる．聞き手が聴取終了した時点で映像を現実時間に復帰させると， RUP 区間に相当する映像フレーム「N5」と「N6」が削除される．この結果生ずる，映像のコマ落ちがこの処理のデメリットである．一方で，コマ落ちで削除されるのは，話し手が聞き手の残余発話を聴取終了するまでを待つ時間である区間は削除した方が適当とも考えられるため会話実験で本手法の評価を実施した．

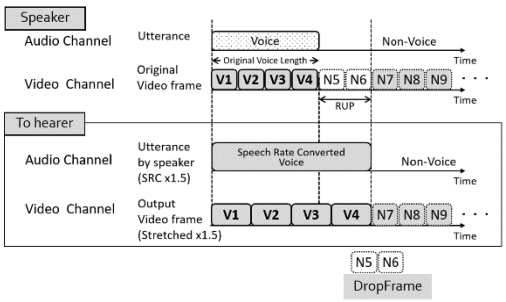

図１　音声と映像が同期する話速変換で生ずるコマ落ち

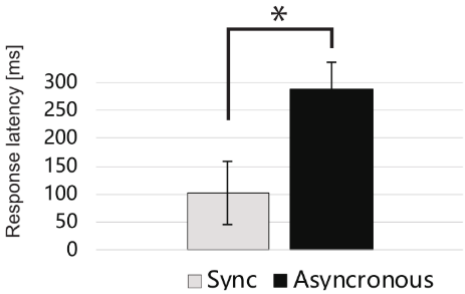

図２　音声と映像が「同期する条件」と「非同期な条件」次話者の平均反応潜時

提案した映像制御手法を実装したシステムにおける 3 人会話実験を収録し，音声・映像非同期条件と，音声・映像同期条件を比較した．話速変換処理を会話に利用した際に，順番交替に関わる認知的負荷が大きいと，次発話開始への反応の遅れ時間が大きくなると考えられる．そのため，提案システムの効果を測る指標として，次話者の反応潜時を計測した．分析する順番交替場面は，「先行話者が非支援対象者（発話に話速変換が適用される参加者)」，かつ「先行話者と次話者が相互に視線を向けている場面」とした．この条件で抽出された場面は，「同期条件」では 61 箇所，「非同期条件」では 62 箇所あった．図２に，それぞれの反応潜時の平均値を示す．反応潜時の平均値は「同期条件」で 101.62ms，「非同期条件」で 288.06 ms となり，F 検定の結果，「同期条件」・「非同期条件」の間で分散の差は認められなかった．これらの結果，話速変換を使った場合であっても映像同期によって，タイムラグを意識することなく次発話を開始可能となり反応潜時が短縮，聞き手にとっては映像同期の正の効果が認められることが分かった．また，実験後のインタビューから，音声・映像非同期では会話のしにくさや，聞き手が認知的負荷を感じていることを確認した．以上の実験を通して，提案する話速変換会話における発話部分の音声と映像同期によって，聞き手は相手の意図を正しく受け取るようになり，次発話をしやすくなったと考える．また，映像同期の効果は次話者指定の意思伝達や，反応潜時の短縮だけに限らず，映像会話で伝達されるジェスチャーなどの非言語情報のタイミングも同時に伝わることで生まれるメリットも示した．この結果は，電子情報通信学会論文誌 A に掲載された[1]．

(2)　話速変換の会話中に生じるタイムラグを補完するユーザインタフェースの策定

達成目標(2)，(3)についての成果を記述する．話速変換会を利用する会話では，話し手の発話終了時刻から，聞き手が伸長された音声を聞き終わるまでに待ち時間（時間差）が生じる．この課題は，ゆっくりとした発話の聴取の支援を必要としない「話し手」に対する支援機能である．

話し手が，聞き手側のゆっくりとした発話の受聴内容を把握できるように，話し手の発話終了後に，話し手自身にループバック音声（音声フィードバック）を提示するインタフェースを実装した．この手法は，先行研究の文献[2]で既に提案していたが，二つの課題が残っていた．一つは，話し手が一方的に発話を継続すると，話し手と聞き手の時間差が増大する．もう一つは，話し手は発話中にボタン操作が必要だったことである．これらの課題に対応するため，本研究期間では，音声区間の自動検出機能を実装し発話中のボタン操作を不要とした．また，話速変換による話し手と聞き手の間のタイムラグの増大を抑

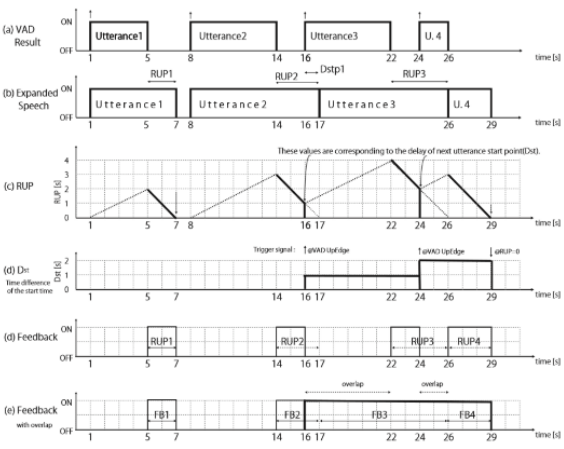

図３　音声フィードバックの生成区間

えるために，聞き手の聴取状況に応じて，音声フィードバックの再生・停止を適応的に切り替えるモデル（図３）を策定し[3]，会話実験により評価をした．その結果，実装した音声フィードバックを聴取する話し手は，音声フィードバックを基準として発話を組み立てる発話様式をとることを明らかにした．また，発話開始のタイミングが一定になることで，参加者間の発話開始時刻の差の変化が小さくなる．それにより，発話話衝突が減少する結果を示した[4]．

（1）斎藤博人，小山内一樹，徳永弘子，武川直樹 ，“ 話速変換を利用する会話における映像同期手法とその効果 ”，信学論 A, Vol.J102-A, No.2, pp.59-67, 2019.
（2）斎藤博人，徳永弘子，橋本恵理子，武川直樹，“リアルタイム話速変換を用いた会話におけるループバックの効果，”信学技報，Vol.115，No.35，pp.67-72，2015.
（3）Ju Hui Peng, Hiroto SAITO，“Implementation of voice feedback model for speaker in speech rate converted conversation”，RISP International Workshop on Nonlinear Circuits, Communications and Signal Processing, pp.671-674, 2018
（4）大場博之，斎藤博人，“話速変換を利用する会話における音声フィードバック生成モデルの検討”，電気学会，電子・情報・システム部門大会，発表論文集 pp.1490-1491, 2019.

５．主な発表論文等

〔雑誌論文〕　計2件（うち査読付論文　2件／うち国際共著　0件／うちオープンアクセス　1件）

１．著者名 斎藤博人，小山内一樹，徳永弘子，武川直樹	４．巻 J102A
２．論文標題 話速変換を利用する会話における映像同期手法とその効果	５．発行年 2019年
３．雑誌名 電子情報通信学会　論文誌A	６．最初と最後の頁 59-67
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 －

１．著者名 斎藤 博人、熊谷 功介、徳永 弘子、武川 直樹	４．巻 J101-D
２．論文標題 話速変換会話における遅れ時間の可視化とその効果	５．発行年 2018年
３．雑誌名 電子情報通信学会論文誌D 情報・システム	６．最初と最後の頁 348～358
掲載論文のDOI（デジタルオブジェクト識別子） 10.14923/transinfj.2017HAP0014	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 －

〔学会発表〕　計5件（うち招待講演　0件／うち国際学会　1件）

１．発表者名 大場博之，斎藤博人
２．発表標題 話速変換を用いた会話における話し手への音声フィードバック手法の検討
３．学会等名 電気学会 電子・情報・システム部門大会
４．発表年 2018年

１．発表者名 Ju Hui Peng, Hiroto SAITO
２．発表標題 Implementation of voice feedback model for speaker in speech rate converted conversation
３．学会等名 International Workshop on Nonlinear Circuits and Signal Processing NCSP18（国際学会）
４．発表年 2018年

1．発表者名 小山内一樹，徳永弘子，武川直樹，斎藤博人
2．発表標題 話速変換会話における音声への映像同期の効果
3．学会等名 電子情報通信学会HCS研究会3月
4．発表年 2018年

1．発表者名 小山内一樹，徳永弘子，武川直樹，斎藤博人
2．発表標題 話速変換会話における映像音声の同期・非同期再生条件の比較検討 ～ 話者が伝える宛先指定の強さは変化するのか ～
3．学会等名 電子情報通信学会HCS研究会5月
4．発表年 2017年

1．発表者名 大場博之，斎藤博人
2．発表標題 話速変換を利用する会話における音声フィードバック生成モデルの検討
3．学会等名 電気学会，電子・情報・システム部門大会
4．発表年 2019年

〔図書〕　計0件

〔産業財産権〕

〔その他〕

―

6．研究組織

	氏名 （ローマ字氏名） （研究者番号）	所属研究機関・部局・職 （機関番号）	備考
研究分担者	武川　直樹 (Mukawa Naoki) (20366397)	東京電機大学・システムデザイン工学部・教授 (32657)	