

科学研究費助成事業 研究成果報告書

平成 29 年 6 月 16 日現在

機関番号：32657

研究種目：基盤研究(C) (一般)

研究期間：2014～2016

課題番号：26330232

研究課題名(和文) 時間管理モデルと感情印象の修正による話者の意図が伝わる話速変換会話システムの実現

研究課題名(英文) Realization of Speech Rate Converted Conversation System Conveying Speaker's Intention by use of Time Management Model and Modifying Emotional Impression

研究代表者

齋藤 博人 (SAITO, Hiroto)

東京電機大学・情報環境学部・准教授

研究者番号：00328519

交付決定額(研究期間全体)：(直接経費) 3,500,000円

研究成果の概要(和文)：話速変換を用いたときの3人会話の分析を行い、参加者らが発話開始のタイミングを把握できる話速変換会話システムを開発した。最初に、話し手が聞き手側の聞き終わりを把握できるように、話者の発話終了後に話者にループバック音声を受聴させる機能を話速変換会話システムに実装した。次に、時間伸長によって生じた待ち時間を画面上に配置したメーターによって可視化し、待ち時間を全参加者が把握できるインタフェースを実装した。会話実験を実施し反応潜時を分析した結果、待ち時間の可視化によって、平均反応潜時が有意に短縮することが明らかとなった。これらは、先行話者と後続話者の円滑な話者交替を可能に効果があることを示した。

研究成果の概要(英文)：The study analyzes three party conversations with speech rate conversion technology, and developed a novel speech rate conversation system that enable the participants to recognize timing to begin their utterances. First, we propose an insertion of speaker's loop-back voice for him/her in speech rate converting conversation system that enable the speaker to recognize hearer's listening end time. Then, we propose and developed a visual feedback functions that can share the listening end time information for facilitating turn-taking among conversation participants. The experimental results show that proposed method was effective for decreasing response latency. It suggests that the proposed method contributes to smoothing turn-takings and also to performance improvements of conversations.

研究分野：信号処理，インタフェース

キーワード：話速変換 順番交替 音声フィードバック 視覚フィードバック

1. 研究開始当初の背景

話速変換技術は、音声の高さを保ちつつ発話速度を変化させる技術である。話速変換によるゆっくり再生は、単に聞き取りを容易にするだけでなく、発話の理解や意図の解釈といった認知的側面も支援し、コミュニケーションに重要な意思疎通に貢献すると考えられる。この技術は、非母語者や高齢者の聞き取りの支援に有効である。しかしながら、話速変換のゆっくり再生には音声遅延を伴うため、会話で利用すると、発話衝突等のトラブルにより順番交替がしにくいなる問題が生じる。そのため、これまでに話速変換技術がリアルタイムの会話に利用される例はほとんど見られず、テレビやラジオのように聴取のみを行う場面の適用に留まっている。また、話速変換は、言語情報の聞き取りやすさを向上させる一方、感情印象を変容させ、参与者同士の意思や感情の理解を阻害する可能性もある。話速変換を会話コミュニケーションに適用する場合は、音声そのものの聞き取りやすさに加え、話者の意図を正確に伝達でき、順番交替のしやすさを実現する技術とそのインタフェースが求められている。

2. 研究の目的

本研究ではリアルタイム話速変換に「順番交替のしやすさ」と「感情印象・意図の修復」を取り入れ、複数人会話においても対等なインタラクションができる話速変換方式の実現を目的とする。

3. 研究の方法

(1) 音声フィードバックによる話し手の話しやすさの支援

話速変換会話での順番交替において、音声遅延の起因する発話衝突や沈黙が起こる問題の解決を目的に、話し手が聞き手側の聞き終わりを把握できるように、話者の発話終了後に話者がループバック音声（話速変換された話者自身の音声）を受聴する話速変換会話システムを実現し分析する。

(2) 視覚フィードバックによる聞き手の応答のしやすさを支援

音声の時間伸長によって生じた音声遅延の長さを画面上に配置したメーターによって可視化し、遅れ時間を全参加者が把握できるインタフェースを構築し分析する。

(3) 次話者指定の意図伝達の評価

映像付き会話システムでの話速変換会話において、音声と映像が同期する会話システムを構築し、映像同期する条件と、非同期の

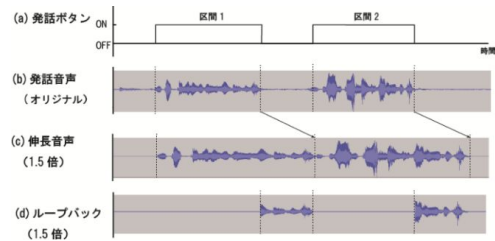


図1 発声音声（オリジナル）の波形と、1.5倍の時間伸長で話速変換処理した波形、及び話し手に呈示されるループバック音声の波形の処理例

条件間で宛先指定の強さに変化が生ずるのかを実験により明らかにする。

4. 研究成果

(1) 音声フィードバックによる話し手の話しやすさの支援

話速変換会話では、順番交替において発話衝突や沈黙が起こる問題があった。これは、話し手側が発話を完了しても、聞き手側が話し手の発話時間長に伸長率から決まる遅れが加わった時刻まで聞き終わることができないため、聞き手の聞き終わりの時刻や次の発話開始の時刻の判断がしにくいためである。この問題の解決を目的とし、話し手が聞き手側の聞き終わりを把握できるように、話し手が聞き手の受聴終了時点を把握できる商法を提案し、その効果を実験により検証した。〈音声フィードバックにより受聴箇所を把握する話速変換会話システムの構築〉

提案法である音声フィードバックによる話し手の支援法について説明する。図1の波形(b)は話者の発話（波形）であり、波形(c)は話速変換処理により時間長1.5倍に変換した波形である。この場面は、両者の発話開始時刻は一致しているが、聞き手が波形(c)を受聴完了する時刻は、波形(b)より遅れていることがわかる。したがって、話者は自分の発話終了後に、聞き手が受聴を完了するまで遅れ時間がある。この遅れ時間は、システムの時間伸長による“話者に生ずる余分な遅れ時間”である。この余分な遅れ時間は、リアルタイムの会話では存在しない時間であるため、話し手が会話に不自然さを感じる原因となる。そこで、この遅れ時間に聞き手が受聴している発話（波形(d)）を話し手にループバックすることで、話し手が感じる不自然さの軽減を狙った。波形(d)をフィードバック音声と呼び、話し手の発話と重複しないように、話し手の休止区間のみでフィードバック音声を提示するシステムを構築した。

映像観測から、ループバック音声を受聴した話し手は、発話内ポーズにおいて、自身の音声を受聴した後に後続発話を開始する発話様式をとることが観測された。話者がこの発話様式で話すことは、発話衝突の回避に効果があるだけでなく、話速変換会話においてもゆっくり発話をする支援対象者による言い差しが可能になることを確認した。これは、話速変換会話で問題となっていた、発話衝突の後の高齢者の言い直しを無くすることにもなり、参加者らのストレスの軽減に寄与する。

(2) 視覚フィードバックによる聞き手の応答のしやすさを支援

時間伸長によって生じた遅れ時間を画面上に配置したメーターによって可視化し、遅れ時間を全参加者が把握できるインタフェースを構築した。遠隔3人会話実験を実施し、反応潜時を分析した結果、遅れ時間の可視化によって、平均反応潜時が有意に短縮することが明らかとなった。以下に詳述する。

<話者交替のしにくさの原因の検討>

話速変換会話における話者交替のしにくさの原因の一つは、交替潜時が長くなる点である。長い交替潜時は、後続話者の本来の意図と異なり、応答をネガティブな意味に変容することを意味している。この変容は、先行話者と後続話者との間に混乱を生み、話者交替をしにくくする。特に先行話者の場合、長い交替潜時により自身の発話が受け入れられなかったと感じたとき、発話を継ぎ足し、図2のような発話衝突のトラブルを誘発していたことが分かった。

<交替潜時を増加させる要因の分析>

話速変換会話において交替潜時が増加する要因は主に2つある。一つは、先行話者の知覚によるものである。通常会話では、先行話者から見た話者交替間隔である交替潜時 (switching pause) と後続話者から見た話者交替間隔である反応潜時 (response latency) の長さは図3の左側のように同一である。一方、話速変換会話では、音声を伸長することでゆっくりな会話を実現しているため、後続話者は伸長された音声を聴取し終えてから発話を開始することになる。そのため、図3の右側のように先行話者は自身の発話終了から後続話者が伸長音声を聴取終了するまでの時間 (遅れ時間) だけ交替潜時を反応潜時より長く知覚することになる。

もう一つの要因は、後続話者の発話開始の遅れである。話速変換会話における遅れ時間の長さは発話時間と伸長倍率から決まり、時間長は一定でない。そのため、IP電話や衛星電話などで生じる遅れと比べ、非常に長い遅れ時間が発生することもある。遅れ時間が発生した際に後続話者が発話を開始すると、相手の発話 (既に発話されメモリ上にスタックしている伸長音声) の再生は中断されず、後

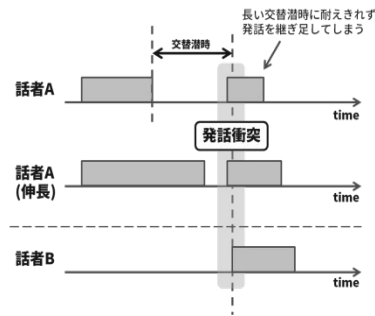


図2 長い交替潜時によるトラブル

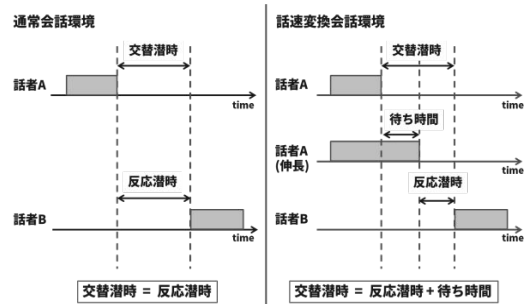


図3 通常会話環境と話速変換会話環境の比較

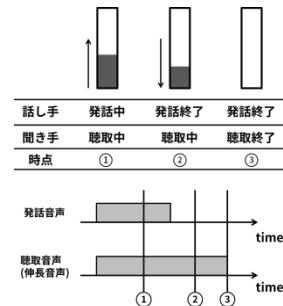


図4 待ち時間とメーター動作の対応関係

続話者の発話が先行話者の発話に対する妨害となるトラブルが発生する。このようなトラブルが生じると、後続話者は話者交替に慎重となり意図的に反応潜時を長くなるとの報告もあり、結果として交替潜時も長くする要因となっていた。

<提案手法：遅れ時間の可視化>

先行話者が反応潜時に比べて交替潜時を長く知覚するのは、反応潜時と遅れ時間を区別せず、遅れ時間を交替潜時に含めたためである。また、後続話者の発話が、本人が意図しないにも関わらず相手の発話に対する妨害となるのは、先行話者の発話にどれだけ遅延があるのか分からずに、発話を開始するからである。つまり、双方に生じた問題の原因は、伸長処理によって生じる遅れ時間を把握できていない点にある。したがって、システムが遅れ時間を提示し把握可能になれば、双方に生じる問題の解決を期待できる。先行話者が、相手が伸長分の音声を聴取している時間を正しく把握できれば、先行話者は遅れ時間を含まない交替潜時を知覚できる。また、後続話者が遅延を把握できれば、遅延がなくなってから話すことで次発話が妨害となる

トラブルを避けることが容易となり、反応潜時を短くできる。

そこで、提案手法では伸長処理によって生じた遅れ時間を可視化し、相手映像に重畳して表示することで先行話者側と後続話者側の双方が遅れ時間を把握可能にする。

提案手法では遅れ時間をメーターの長さによって可視化した。メーターの動作と遅れ時間の関係について図4を用いて説明する。

メーターの上昇(図4:)

話者が発話中: 発話に伸長処理が適用され遅れ時間が増加している状態である。

メーターの下降(図4:)

話者が発話を終了: 伸長処理が止まり遅れ時間が減少している状態である。この状態での参加者の発話は会話への妨害となる。

メーター残量無し(図4:)

遅れ時間が0: 伸長音声の再生が完了し、聞き手の聴取が終了して話し手と聞き手の間に時間差が無い状態である。この状態での話者交替であれば、後続話者の発話開始が時間差によって先行話者の発話への妨害となるトラブルは発生しない。

<画面インタフェースの検討>

3人の話速変換会話において、メーターにより遅れ時間を可視化した様子を図5に示す。遅れ時間には、自身の発話によって生じるものと相手の発話によって生じるものの2種類があるため、両者の遅れ時間は独立に提示する。インタフェースは、自身の発話により生じる遅れ時間を画面の内側に配置し、相手の発話によって生じる遅れ時間を画面外側に配置した。なお、発話に話速変換を適用しない場合は遅れ時間が発生しないため、話速変換を適用しない話者のメーターは表示しないよう設定した。

<実験の概要>

遅れ時間の可視化が話速変換会話の話者交替に対し、どのような効果を及ぼしたのかを分析するため、会話実験を実施した。実験には、友人同士の学部学生3人組2ペア(計6名)が参加した。各ペア内からは話速変換による支援を受ける対象を1名とし、他の2名の発話に話速変換処理を行う。図5に示す聴取環境で会話を行った。会話は10分間の自由会話とし、以下の3条件の会話を収録した。

条件1: 話速変換会話

条件2: 話速変換会話(話者への遅れ時間の可視化)

条件3: 話速変換会話(話者と相手への遅れ時間の可視化)

可視化無しの実験条件1と可視化有りの実験条件3の会話後半5分間の分析を行った結果、分析対象となった反応潜時は119個、遅れ時間が発生しない先行発話が原音声である場合を除くと計90個であった。両条件の反応潜時の平均および標準偏差SDを表1、平均反応潜時のグラフを図6に示す。標準偏差が大きく異なっており分散の大き



図5 画面インタフェース(話者A視点)

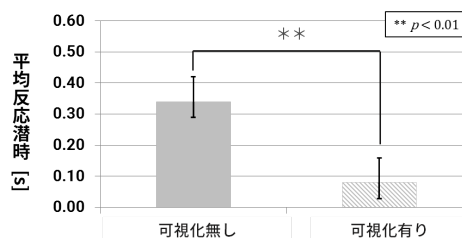


図6 平均反応潜時

表1 反応潜時の分析結果

	without- visualization	visualization
N	39	51
\bar{X}	0.34	0.08
SD	0.50	0.37

さが等質でないと考え検定を行った結果、両条件の分散の差は有意であった(両側検定: $F(39,51)=1.86, p<0.05$)。そのため、等分散を仮定しないウェルチの法によるt検定をおこなった。その結果、両条件の平均の差は有意であった(両側検定: $t(67)=2.75, p<0.01$)。したがって、可視化有り条件は、可視化無し条件に比べて平均反応潜時が短くなるといえる。提案手法を用いることで話速変換会話における平均反応潜時が有意に短縮することが分かった。話速変換会話における交替潜時は反応潜時と遅れ時間の和であることから、提案手法は交替潜時の短縮としても有効である。

遅れ時間の可視化は、先行話者と後続話者の円滑な話者交替を可能にし、会話のしやすさに寄与する。

(3) 次話者指定の意図伝達の評価

リアルタイムに双方向のやり取りが生じる会話においては、現在の映像情報を伝送した方が話者交代の手掛かりとなる情報を活用しやすいという考えのもと、映像には伸長加工を施さずに伝送していた。そのため、ゆっくり音声を受聴している区間は、映像に対し音声が遅れ、会話者らはリップシンクしない映像を見ていた。しかし、人間同士の会話では、視線行動やジェスチャーといった視覚により伝達される情報もあり、その中には話

し手の次話者指定に関する行動も含まれる。この映像と音声のズレが会話における意思の伝達に支障をきたし、話し手の意図が別な意味に解釈されることがあった。

話速変換音声に映像が同期する会話システムを構築し、映像同期・非同期の条件間で宛先指定の強さに変化が生ずるのかを実験により検証した結果、特に話し手が発話未付近に次話者指定行動を表出している場合には、宛先指定の強度が非同期条件よりも強く伝達されることを明らかにした。

(4) 成果のまとめ

本研究は、話速変換会話で生ずる話しにくさを解決するために、話者の発話終了後に話者にループバック音声（話速変換された話者自身の音声）を受聴させる機能を有する話速変換会話システムを提案し、会話実験とその分析から、音声フィードバックは、発話衝突の回避に効果があるだけでなく、ゆっくり発話をする支援対象者による言い差しが可能になることを示し、話速変換会話における有効性を示した。

また、遅れ時間をメーターによって可視化する可視化する会話システムを提案し、会話実験の結果、提案手法を用いることで話速変換会話における平均反応潜時が有意に短縮することが示された。話速変換会話における交替潜時は反応潜時と遅れ時間の加算であることから、提案手法は交替潜時の短縮としても有効であるといえる。提案手法によって実現した順番交替のタイミングまでの時間を把握する支援機能には、話速変換会話における会話のしやすさに大きく貢献できると期待できる。今後は、音声フィードバック、視覚フィードバック、映像同期手法を統合し弱者（非母語話者や認知機能が衰えた高齢者）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について明らかにしていく。

5. 主な発表論文等

〔学会発表〕(計 11 件)

熊谷功介, 徳永弘子, 武川直樹, 齋藤博人, “話速変換会話における参与者の順番交替のデザイン ~ 待ち時間の可視化が後続話者の発話に与える影響 ~”, 電子情報通信学会 HCS 研究会, Vol.116, No.524, pp.155-160, 2017/3/16, 東北大学(宮城県・仙台市)
齋藤博人, 徳永弘子, 武川直樹, 大島直樹, “遠隔地家族間のコミュニケーションを引きだす食卓メディアの構築 ~ 料理中の動作状況を推定する信号解析法の検討 ~”, 電子情報通信学会 CNR 研究会, Vol.116, No.461, pp.171-174, 2017/2/19, 北海道大学(北海道・札幌市)

齋藤博人, “話速変換を会話に利用するためのインタフェースデザイン”, 電子情報通信学会 HC 第 55 回 VNV 研究会, 2016/12/26, 国立情報学研究所(東京都・千代田区)

熊谷功介, 徳永弘子, 武川直樹, 齋藤博人, “リアルタイム話速変換会話における聴取時間可視化のシステムデザインと実験評価”, 電子情報通信学会 HCS 研究会, Vol.116, No.31, pp.145-150, 2016/5/19, 沖縄産業支援センター(沖縄県・那覇市)
廣瀬貴大, 熊谷功介, 徳永弘子, 齋藤博人, 武川直樹, “フィードバック音声を聴取する話速変換会話における話者の宛先指定行動分析”, 電子情報通信学会技術研究報告書 Vol.115, No.35, pp.67-72, 2016/1/23, やまと会議室(奈良県・奈良市)

川端啓太, 齋藤博人, 武川直樹, “「うん」と「はい」が応答されるまでの時間が発話の印象に与える影響”, 電子情報通信学会技術研究報告書 Vol.115, No.418, pp.107-112, 2016/1/23, やまと会議室(奈良県・奈良市)

齋藤博人, 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音声ループバックの効果”, 電子情報通信学会技術研究報告書 Vol.115, No.35, pp.67-72, 2015/5/19, 沖縄産業支援センター(沖縄県・那覇市)

小野貴大, 齋藤博人, 金子博, 武川直樹, “音声パラメータと感情印象変化の対応モデルの構築 話速とピッチを制御された音声の感情印象の改善”, 電子情報通信学会技術研究報告書 Vol.114, No.440, pp.193-198, 2015/1/30, ベイリゾート小豆島(香川県・小豆島)

橋本恵理子, 齋藤博人, 徳永弘子, 武川直樹, “遠隔会話でのリアルタイム話速変換が高齢者ユーザの行動に与える影響の分析”, HCG2014-1-1-8, pp.149-155, 2014/12/17, 海峡メッセ下関(山口県・下関市)

〔その他〕

<http://www.sps.sie.dendai.ac.jp/>

6. 研究組織

(1) 研究代表者

齋藤博人 (SAITO, Hiroto)
東京電機大学・情報環境学部・准教授
研究者番号: 0 0 3 2 8 5 1 9

(2) 研究分担者

武川直樹 (MUKAWA, Naoki)
東京電機大学・情報環境学部・教授
研究者番号: 2 0 3 6 6 3 9 7