

令和 6 年 6 月 14 日現在

機関番号：32657

研究種目：基盤研究(C) (一般)

研究期間：2020～2023

課題番号：20K11922

研究課題名(和文) 時間伸長を知覚せずにゆっくり音声を聴取する公平な会話機会を有する会話環境の実現

研究課題名(英文) Realizing Equitable Conversational Environments: Facilitating Slow Speech Perception Without Time Dilation

研究代表者

齋藤 博人 (Saito, Hiroto)

東京電機大学・システムデザイン工学部・教授

研究者番号：00328519

交付決定額(研究期間全体)：(直接経費) 3,300,000円

研究成果の概要(和文)：話速変換によるゆっくりとした再生には、発話の伝達に音声遅延が発生し、これが会話の円滑な進行を妨げる問題があった。本研究では話速変換を用いる会話において、話者の発話待ち時間を支援する方法を確立し、話速変換を会話で実用化することを目指した。具体的には、次の発話を開始するタイミングを把握するための「視覚フィードバック」と「聴覚フィードバック」機能をリアルタイム会話システムに実装し、その効果と特性を定量的かつ定性的に評価した。その結果、「視覚フィードバック」は発話権が均等な雑談や討論の場面での有効であり、「聴覚フィードバック」は説明形式の会話において有効なフィードバック手法であることが明らかとなった。

研究成果の学術的意義や社会的意義

話速変換会話における主たる支援対象者(高齢者、非母語の学習者)には、周囲に遠慮すること無く理解しやすい音声を提供した上で、健常者(若年者、母国話者)には、遅延を違和感なく許容できるインタフェースとして設計している。健常者には、会話の様式により、異なるフィードバック方式(視覚、聴覚)を選択している。これまでのシステムはユーザ全員に性能向上させることを目指す開発が多いのに対して、本研究では、支援対象者と健常者の対等なコミュニケーションが実現をするものであり、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点に意義がある。

研究成果の概要(英文)：Speech rate conversion technology involves audio delays proportional to the length of the utterance, leading to difficulties in conversational exchanges. This study focuses on the waiting time experienced by the speaker in conversations using speech rate conversion (i.e., the time until the listener finishes hearing the slowed speech). Specifically, we quantitatively and qualitatively evaluated the characteristics of "visual feedback" and "auditory feedback" functions, which help the speaker determine the timing for the next utterance in different conversation styles. As a result, we found that "visual feedback" is effective in scenarios such as casual conversations and discussions where speaking turns are equally distributed, while "auditory feedback" is an effective method in explanatory conversation formats.

研究分野：信号処理，インタフェース

キーワード：話速変換 会話分析 フィードバック

### 1. 研究開始当初の背景

非母語で会話をする学習者や認知機能が衰えた高齢者にとって、相手の発話スピードについていくことはやさしいことではなく、会話の流れから取り残されることがある。こうした聞き取りに対する支援には話速変換技術が有効である。話速変換により聴取音声はゆっくりとなる特性は、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などの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。

しかし、話速変換によるゆっくりとした再生には発話長に比例する音声遅延が伴うため、会話のやりとりがしにくくなる問題があった。リアルタイムのオンライン 2 人会話に話速変換処理を適用すると、1 人は、ゆっくりした速度に伸長された自身の発話がフィードバックされる者(以降参加者 A と呼ぶ)と参加者 A のゆっくりした速度の音声を受聴する者(聞き取りに困難を生じて本来サポートの対象となる者で以降、参加者 B と呼ぶ)の間には、通常のオンライン会話には存在しない時間差が生ずる。図 1 を用いて話速変換による生ずる時間差について説明する。通常のオンライン会話では、交替潜時 (switching pause : 自らが発話を終了してから後続話者が発話開始するまでの時間長) と反応潜時 (response latency : 先行話者が発話を終了してから自らが発話開始するまでの時間長) の長さは図 1(a) のように同一である (ただし、通信経路の遅延は考慮しないものとする)。

一方、話速変換を用いる会話では、図 1(b) に示すように、先行話者 (参加者 A) に認知される交替潜時は、伸長音声を聴取している後続話者 (参加者 B) の反応潜時より長くなる。これは、話速変換の特徴であるが、複数人の会話に適用した場合においては、話し手が次々に入れ替わる順番交替で、発話の重複や不自然な間の要因になり会話がしにくくなる問題があった[1]。

こうした問題に対処するため、我々は聞き手が聴取している音声を話し手にも聞かせる聴覚フィードバック技術[2]と、聞き手の発話聴取時間長を話し手にメーターで提示する視覚フィードバック技術を開発した[3]。それぞれのシステムを実装した会話環境を構築し、被験者による評価実験を行ったところ、話し手が話し続ける際の行動に相違があることが確認された。具体的には、聴覚フィードバックを使用した場合には自身の音声を受聴した後に後続発話を開始する[1]のに対し、視覚フィードバックでは聞き手の発話聴取終了前に話し始める傾向が認められた。これらの結果から、フィードバック手法の違いにより人の発話行動が変化する知見を得たものの、発話行動の特性が雑談形式 (参加者らが自由に発言する対話)、あるいは説明形式 (一方が主な話し手となりもう一方が聞き手になる対話) といった異なる会話形式に実装した場合の影響は明らかになっていない。

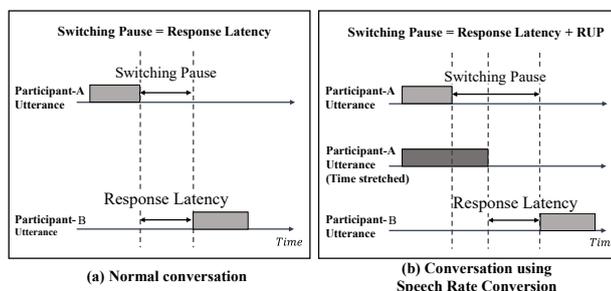


図 1 通常の会話環境と話速変換を用いた会話環境の交替潜時と反応潜時

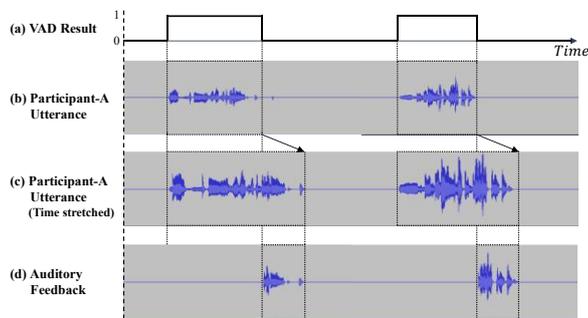


図 2 発声音の波形と、話速変換処理した波形、及び聴覚フィードバックの処理例

### 2. 研究の目的

話速変換を利用する会話の場で、支援対象者には「自らが支援を受けていると感じさせることなく、あたかも話し手が実際にゆっくりと発話をしていると知覚できるインタフェース」を実現しつつ、ゆっくり再生が必要無いユーザには「遅延を違和感なく許容できる」会話環境の実現を目的とする。

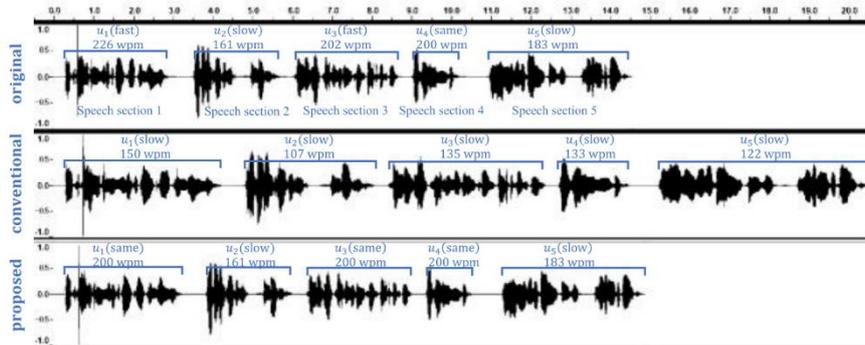


図3 伸長率一定方式と伸長率変動方式で話速変換処理をした結果

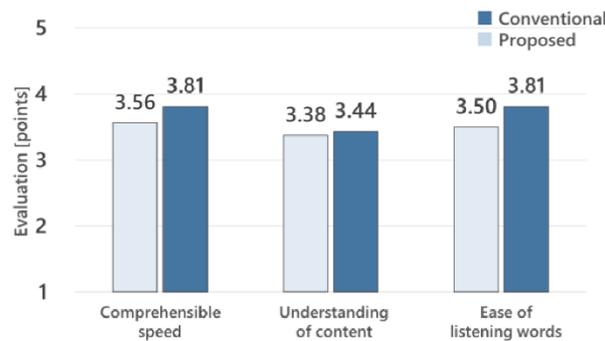


図4 「聞き取りやすい速さ」「内容理解」「単語聞き取り」の評価結果

### 3. 研究の方法

研究は以下のステップで実施した。

(1) ゆっくりした再生が必要な支援対象者向けには、音声と映像がゆっくりになって視聴でき、さらに音声と映像を必要なところだけをゆっくりとした再生にするインターフェースを実現する。直近の発話内容の発話速度から、支援対象者が聞き取りにくいと感じる音声区間を予測して、ゆっくりとした再生が必要な発話を自動選別し再生するモデルを作成する。

(2) ゆっくり再生が必要無いユーザには、雑談形式と説明形式の対話場面を実験的に設計し、「視覚フィードバック」と「聴覚フィードバック」を使用した場合の話し手の対話行動を定量的に分析する。その知見より、会話の様式に応じたフィードバック方式の選択モデルを作る。

### 4. 研究成果

(1) ゆっくりになった音声と映像が同期する Web ブラウザ会話インターフェース、ならびに会話する会話中に発話速度が速くなった場合にだけゆっくりとした再生にするインターフェースの設計および実装をした。この方式の有効性を評価するための音声聴取実験を実施した。実験条件は、従来手法(伸長率一定方式)および、ユーザの好みの速さで聴取する提案手法(伸長率変動方式)の 2 条件である。図 3 にユーザの好みの速さが 200[wpm]の音声の処理結果を示す。提案法では、好みの速さよりも早い音声にのみ話速変換をするため、会話に要する時間を従来法よりも短くできる利点がある。被験者は、聞き取りやすさに関する 3 項目(「聞き取りやすい速さ」「内容理解」「単語聞き取り」)の評価した。その結果を図 4 に示す。

評価結果は、提案方式の評価が従来方式を高く、聞き手にとって理解できる速さで音声を再生しながら、会話に要する時間を低減する効果を示した。

(2) 話速変換を用いる対話において、話し手が、次の発話開始のタイミングを把握するための視覚フィードバック(視覚条件)と聴覚フィードバック(聴覚条件)の特質を比較分析した。

分析は「雑談形式」と「説明形式」の 2 つの対話について、それぞれ視覚条件と聴覚条件での話し手を定量的、定性的に評価した。日本語を母語とする健常な 20 代男女が実験に協力した。共同想起対話は 10 名 5 組(男性 6 名、女性 4 名)であった。協力者らは参与者 A と B の両立場を交替して参加したため、聴覚条件 10 対話、視覚条件 10 対話が収録された。対話は、親しい関

係の同性同士 2 人がペアになった。地図課題対話は 16 名 8 組（男性 8 名，女性 8 名）が参加し，聴覚条件 16 対話，視覚条件 16 対話が収録された。各条件の実施順はカウンターバランスをとった。共同想起対話は聴覚・視覚条件共に 1 対話約 10 分程度の 10 対話であり聴覚条件の対話合計時間は 100 分，視覚条件は 101 分であった。地図課題対話は両条件共に 1 対話約 15 分～16 分程度の 16 対話であり聴覚条件の対話合計時間は 241 分，視覚条件は 263 分が収録された。

収録した会話データを分析した結果，「雑談形式」の対話では「視覚条件」において，メーターを用いて後続発話の開始タイミングが話速変換によって生じた待ち時間（RUP）の終了前後に有意に集中した（図 5，表 1）。その結果，ユーザから「聴覚条件」に比べてタイミングよく話せたと評価された。視覚フィードバックは，発話権が均等な雑談や討論場面への適用が有効であると示唆された。一方，「説明形式」では「聴覚条件」の話し手の後続発話の開始タイミングが「視覚条件」より有意に遅くなった。これは，ユーザからのインタビューから，話し手が聞き手の反応を待ち，共有理解が確認されるまで発話を遅らせる戦略によるものと解釈できた。これは，対話中の相互理解と協調を促進する上で重要と考えられ，診療場面での説明やプレゼンテーションに有効であることが示唆された。今後は，多人数会話への適用，グループディスカッションなど他の会話形式への適用，高齢者，非母語者など本システムのターゲットユーザへの適用に向け，実験に取り組み，得られた知見をシステムのデザインに反映する予定である。

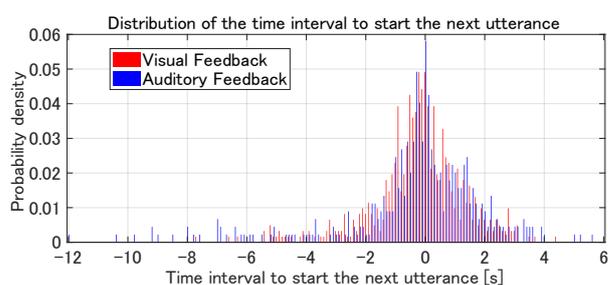


図 5 参与者 A が後続発話を開始タイミングの分布

表 1 参与者 A の後続発話開始タイミングの統計量

Statistics	Visual	Auditory	Significant difference
Median [s]	-0.09	0.01	<i>n. s.</i>
SE [s]	1.55	2.46	<i>p</i> < .01

#### (引用文献)

- [1] 橋本恵理子，斎藤博人，徳永弘子，武川直樹，“発話時間長によって生ずる沈黙が会話の順番交替に与える影響，” SLUD 言語・音声理解と対話処理研究会, Vol.69, pp.67-72, 2013.
- [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，” 電子情報通信学会技術研究報告, Vol.115, No.35, pp.67-72, 2015.
- [3] 斎藤博人，熊谷功介，徳永弘子，武川直樹，“話速変換会話における遅れ時間の可視化とその効果，” 電子情報通信学会論文誌 D, Vol.J101-D, No.2, pp.348-358, 2018.

5. 主な発表論文等

〔雑誌論文〕 計0件

〔学会発表〕 計9件（うち招待講演 0件 / うち国際学会 3件）

1. 発表者名 Kazuma Ban, Hiroto Saito
2. 発表標題 Automatic Switching between Visual Feedback and Auditory Feedback in Speech Rate Converted Conversation
3. 学会等名 NCSP'24 (国際学会)
4. 発表年 2024年

1. 発表者名 Miou Oyama, Hiroto Saito
2. 発表標題 Speech rate Conversion Method with Variable Stretch Rate
3. 学会等名 NCSP'24 (国際学会)
4. 発表年 2024年

1. 発表者名 石川雄大, 齋藤博人
2. 発表標題 話速変換を用いた会話における発話者への聴覚フィードバック方式の検討・聞き手の聴取完了時点を把握する呈示音の評価
3. 学会等名 電子情報通信学会 HCGシンポジウム
4. 発表年 2023年

1. 発表者名 大山美桜, 齋藤博人
2. 発表標題 聴取する音声の速さが一定になる話速変換処理における音声の聞き取りやすさの評価
3. 学会等名 電子情報学会 HCS研究会
4. 発表年 2023年

1. 発表者名 伴一馬, 徳永弘子, 武川直樹, 斎藤博人
2. 発表標題 話速変換を用いる会話における話し手へのフィードバック方式の検討 ~ 音声・視覚フィードバックの適応的な切り替えのための比較評価 ~
3. 学会等名 ヒューマンコミュニケーション基礎研究会
4. 発表年 2022年

1. 発表者名 大山美桜, 斎藤博人
2. 発表標題 会話中に発話伸長率が変化する適応型話速変換システムの検討 発話速度推定精度の改善手法
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 2022年

1. 発表者名 Tamami Mizuta, Hiroto Saito
2. 発表標題 Evaluation of Feedback Methods for Speakers in Speech Rate Converted Conversation
3. 学会等名 RISP International Workshop on Nonlinear Circuits, Communications and Signal Processing 2022 (国際学会)
4. 発表年 2022年

1. 発表者名 水田瑠美, 徳永弘子, 武川直樹, 斎藤博人
2. 発表標題 話速変換された会話における話し手へのフィードバック手法の評価
3. 学会等名 電子情報通信学会 ヒューマンコミュニケーション基礎研究会
4. 発表年 2022年

1. 発表者名 大場 博之, 水田 瑶美, 徳永 弘子, 武川 直樹, 斎藤 博人
2. 発表標題 聞き手の聴取音声速度を一定にする話速変換を利用した会話での話し手への影響
3. 学会等名 電子情報通信学会 HCGシンポジウム
4. 発表年 2020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・研究員    (32657)	

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------