

オンライン会議において相手に与える印象を変化させる 語尾ピッチ変換システム

西原宗太郎^{1,a)} 渡邊拓貴^{2,b)} 寺田 努^{1,c)} 塚本昌彦^{1,d)}

概要: オンラインでのコミュニケーションは話し手の感情、雰囲気といった、音声と映像のみでは伝達が難しい情報が存在する。話し手の心理状態を理解する手段として、音量の強弱や声の高さといった周辺言語が重要視されているが、オンライン会議においてはそれが伝わりづらい。周辺言語の中でも特に語尾のピッチ変化は話し手の性格と印象を決定する重要な要素であるため、オンライン会議において聞き手に与える印象を変化させるために、語尾のピッチを変化させることが有効である可能性が高い。そこで本研究では、オンライン会議において話し手の意図する印象を聞き手に与えるためのオンライン会議システムを提案する。語尾のピッチ変化の有効性を確認するため、提案システムの使用状況を再現した実験を行った。語尾ピッチを上昇させた場合、聞き手は話し手に対して元気である印象を抱く傾向が有意に増加し、提案システムが印象変化に対して有効であることが示唆された。

1. はじめに

新型コロナウイルス感染症の流行により人と人との直接的な接触が制限され、オンライン会議システムが普及しつつある。オンライン会議では対面でのコミュニケーションと異なり、マイクやカメラを通じた音声と映像のみによるコミュニケーションとなるため、話し手の感情や雰囲気を伝達することが困難である。

したがって、オンライン会議においては映像や音声でいかに感情や雰囲気を伝えるかが重要になる。本研究では特に音声情報に着目する。音声によるコミュニケーションは、話の内容そのものを指す言語情報と発話音声のアクセントや音量の強弱、声の高さ、間、発話の速さといった発言内容以外の情報を指す周辺言語の2つの情報で構成される。周辺言語は具体的な事項を伝達する言語情報と違い、話者の感情、雰囲気、態度といった言語化の難しい情報を伝達する。対話中に話し手の周辺言語によって表れる、怒っている、楽しんでいるなどの感情と言語情報の内容が矛盾した場合、聞き手は周辺言語によって表される情報を93%の確率で信頼すると報告されており、周辺言語の聞き手に与える影響は大きい [1]。

また対人コミュニケーションでは、周辺言語を柔軟に変化させることで話の要点の強調や、発言権の自然なやり取りをしている [2]。周辺言語の中でも語尾の音の高さを指す語尾ピッチは話し手の性格と印象を聞き手に類推させる上で重要な役割を果たす [3], [4]。そのため、聞き手に与える印象を良いものにするために、話し手の語尾ピッチを適切に変化させることは有用であると考えられる。

そこで本研究では、オンライン会議においてユーザが相手に与えたい印象を設定するとシステムがその設定に基づいてユーザ発話の語尾ピッチを自動で補正し、ユーザは普通に話すだけで意図する印象を聞き手に与えられるオンライン会議向け語尾ピッチ変換システムを提案する。提案システムはオンライン会議時の発話の語尾を検出し、語尾のピッチを設定に従って変化させる。提案手法の有効性を評価するため、提案システムの使用状況を再現した映像を被験者に視聴してもらい、ユーザから受ける印象を評価する。また、1年間にわたる筆者のシステム利用経験から、システムについて考察する。

2. 関連研究

2.1 周辺言語の役割

コミュニケーションにおける周辺言語の影響を調査する研究が行われている。Banzigerらは、様々な感情が含まれる音声の定量的な韻律分析を行い、発言に含まれる基本周波数の振れ幅、平均を調べることで発言にどのような感情が含まれるかを分類できると述べている [5]。鶴谷は、録音

¹ 神戸大学大学院工学研究科

² 北海道大学大学院情報科学研究科

a) soutarou-nishihara@stu.kobe-u.ac.jp

b) hiroki.watanabe@ist.hokudai.ac.jp

c) tsutomu@eedept.kobe-u.ac.jp

d) tuka@kobe-u.ac.jp

音声の分析を行うことにより、日本語における丁寧な発話に現れる韻律的特徴として、ゆっくり話す、低めに話す、語尾のピッチを次第に上げる傾向にあると述べている [6]. Dahlback らは、被験者の居住地域特有のアクセントとそれとは異なるアクセントの2種類の英語を使用した場合には前者の方に類似性を感じ、信頼する傾向にあると評価した [7]. 市川らは、非対面で行われた会話データに含まれる声の大きさ、高さ、間、発話の早さからなる抑揚情報を解析することにより、会話相手の抑揚情報に合わせて自身の抑揚情報を操作することで発言権、話題の転換のやり取りを制御していると述べている [2]. 郡は、日本語における語尾のアクセントを分類し、それらを再現した合成音声聞いた際に受ける印象の調査を行い、合成音声でも語尾のアクセントの違いにより聞き手の受ける感情が異なることを確認した [3]. これらの先行研究から、オンライン通話での入力音声に補正を施すことにより聞き手が受ける印象に影響を与える可能性がある。発話に含まれる印象に応じて周辺言語を変化させるシステムは、聞き手が受ける印象を変えるために有用であるが、オンライン通話上でも同様の効果が得られるかを調査した研究は、筆者の知る限り存在しない。

2.2 音声加工による補助

音声信号処理技術の発達により、ソフトウェアなどを用いて発話音声のピッチやイントネーションを変化させることが可能になっている。常島らは、発話における助詞や語尾のピッチ変化によって聞き手が受ける印象に影響を与えるかについて調査を行い、語尾と助詞の基本周波数を下降させた音声は通常の音声と比べて、「経験が豊富である」、「説得力がある」などの印象をより強く与えることを示した [8]. しかし、この研究では、あらかじめ録音した音声の語尾と助詞を加工したものを使用しており、日常のコミュニケーションへ応用することを想定するとリアルタイムで加工されることが望ましい。亀岡は、音声を構成する基本周波数を作り出す役割をもつ器官である甲状軟骨のメカニズムを模擬した制御物理モデルによる基本周波数パターンの高精度な推定を行うことにより、発話者特有の基本周波数を求め、入力音声のアクセントを自由に操作することを可能にした [9]. 外川らは、携帯電話での受信音と周囲雑音のSN比に応じた利得の制御による音声強調技術、無音区間の短縮と声質を変えない音声の引き延ばし技術による話速変換技術を用いて聞こえやすさ支援をした [10]. この研究では、発話音声全体を加工することで、発話内容の伝達を補助する技術が実装されているが、発話音声加工することによる印象表現を制御するものではない。

3. 提案システム

本研究では、聞き手により良い印象を与えるために、語

表 1: 語尾のアクセントと得られる効果の関係

	効果	アクセント
強調型上昇調	ぜひわからせたい気持ちを込める	分かり ます
急下降調	気づいて分かったことを示す	分かり ます
疑問型上昇調	答えを求める・反応を待つ	分かり ます
上昇下降調	気づかせ、わからせたい気持ちを込める	分かり ます

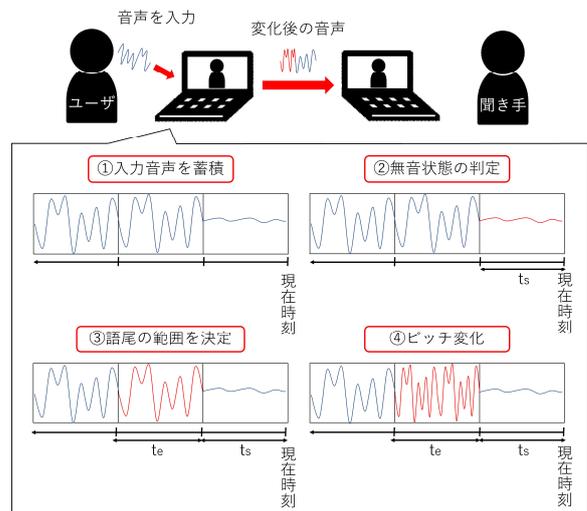


図 1: システムの流れ

尾ピッチをユーザが設定した高さに自動に変化させるシステムを構築することを目指す。

3.1 システムの概要

提案システムは語尾ピッチを変化させることで、相手に与える印象を変化させるものである。語尾のアクセントの種類と得られる効果をまとめたものを表 1 に示す。表 1 は 2 章の関連研究で挙げた複数の関連研究から得られた語尾変化の印象特徴を示したものであり、同じ発言内容であってもアクセントの違いにより相手を受取る印象が異なることが分かる。提案システムでは、入力された音声に対して語尾のアクセントを与えたい印象に合わせて変更するピッチ変化を行う。提案システムの処理の流れを図 1 に示す。システムは、ユーザの発話をマイクで取得、蓄積し、ユーザの発話から語尾を自動検出した後、ユーザの意図に沿ったピッチ変化を語尾に適用し、聞き手側の PC に送信する。

3.2 語尾ピッチ変化手法

会話中に語尾ピッチを変化させる場合、リアルタイムでマイク入力を会議システムに出力すると、ピッチ変化を行う前に音声出力されてしまう。よって提案手法では、マイク入力を一定期間保持し、発話を終えたことを認識してから語尾にあたる部分に遡ってピッチ変化を適用し、その後会議システムへ音声出力する必要がある。発話終了タ

イミングの検出には、発話終了後の無音区間を用いる。無音状態が一定時間続いた時の無音区間の開始時点が発話終了タイミングとする。発話終了タイミングから語尾の長さだけ遡った範囲を語尾として、ピッチ変化を適用する。図1では、語尾にあたる部分の長さを t_e 、発話終了後の無音区間を t_s としている。 t_e と t_s の長さを定義する必要があるが、 t_e と t_s の長さは使用する語尾の種類、話す速度によって変化するため一意に定めることは困難である。したがって本研究では、予備実験より t_e と t_s の長さを定めた。 t_s の長さの設定は、発言中に文節で区切ったり息継ぎしたりすることで音声途切れる瞬間との誤認識を防ぐための適切な値にする必要がある。1対1で行われた1分程度の会話を4種類録音し、録音音声の分析を行った結果、考え込むなどの無音区間を除いた、会話に含まれる短い無音区間は平均0.53秒であった。この値にマージンをもたせ、本研究では、0.7秒間を t_s の長さとして定義する。提案システムでは、入力音声信号が60dB以下であれば無音としている。録音音声の語尾の長さの平均は0.23秒であった。分析を行った録音音声には「です」、「ます」、「よ」、「ね」、「か」といった語尾が含まれていた。語尾区間では語尾開始時から次第に音量が小さくなっていくため、語尾終了時付近の音声を除いた0.2秒を t_e の長さとして定義する。したがって、0.7秒間の無音区間があった場合、直前の0.2秒間の音声データを語尾としてピッチ変化の対象とする。 t_e と t_s の長さの合計0.9秒間の音声を蓄積し、蓄積した音声にピッチ変化を施した音声を聞き手側へ出力する。そのため、提案システムの出力音声は0.9秒の遅延が生じる。無音状態が0.7秒以上続かなければ蓄積した音声を取り出してピッチ変化を行わず聞き手側へ出力し、0.7秒以上の無音状態を確認した際には、 t_e にあたる範囲のみをピッチ変化した音声を聞き手側へ出力する。提案システムでは表1で示した、強調型上昇調、急下降調、疑問型上昇調、上昇下降調 [3] をピッチ変化により実現する。

3.3 実装

入力音声の語尾ピッチ変化を行うシステムのプロトタイプを実装した。実装はPythonを用いる。出力音声は仮想ミキサ VoiceMeeter Banana を用いて聞き手に送信する。実装したシステムのGUIを図2に示す。図2の画面左下はPCに入力される音声のピッチを表示し、画面右下は聞き手側へ出力する音声のピッチを表示している。実験では試験的に語尾のピッチを上昇させる場合は語尾に含まれる基本周波数の1.2倍の周波数に、語尾のピッチを下降させる場合は語尾に含まれる基本周波数の0.8倍の周波数にしたが、GUIによりその設定は自由に変更できるようになっている。ピッチ変化はTD-PSOLA方式 [11] を用いた。また、提案システムの使用により音声は0.9秒遅延するため、映像出力も音声に合わせて0.9秒遅延させて出力した。こ

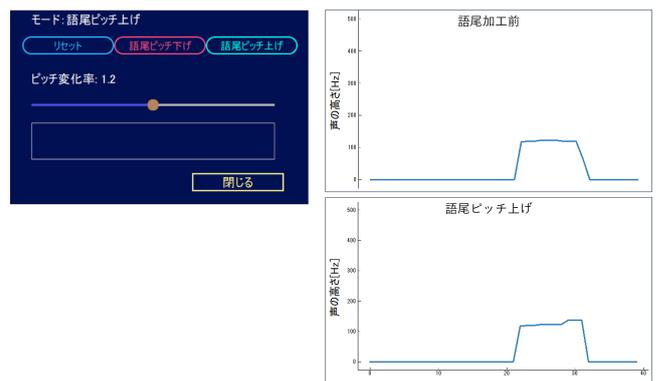


図 2: システムの GUI

の処理には Open Broadcaster Software の仮想カメラ機能を用いた。

4. 評価実験

提案システムの評価を行うために、まず、オンライン会議において提案システムを利用した際に、聞き手が受ける印象の変化について評価し、次に、1年にわたるシステムの実利用の結果を考察する。

4.1 実験の設計

本稿では、オンライン会議における語尾ピッチ補正が、ユーザの印象を変化させるのかをまず明らかにするため、語尾ピッチ補正を行った会話動画を用いた評価実験を行った。提案システムによって検出される語尾の位置が、実際の語尾の位置とずれが生じると、語尾ピッチ補正による影響のみを評価できないため、提案システムが理想的に動いていることを確認した動画を実験に用いた。オンライン会議において聞き手の受ける印象に影響を与えたいような状況を想定し、1対1で会話を行った動画を用意する。ユーザと聞き手が対話している様子を被験者が動画で視聴し、システムを利用した場合と利用しない場合とで被験者がユーザに対して受ける印象についてそれぞれ主観評価を行う。対話において聞き手が受ける印象のうち、提案システムの音声補正が影響を与えると考えられる印象は「自信がある」「説得力がある」「元気がある」「丁寧である」である。先行研究より、対人コミュニケーションにおいて語尾ピッチを下降させることで「丁寧さ」「言い切り」の印象が強くなることが報告されており、「言い切り」は説得力の向上につながると考えられる [6], [12]。また、語尾ピッチを上昇させることで「力強さ」の伝達が可能になる [13]。「力強さ」を伝達することは元気、自信があることにつながると考えられる。したがって、本研究ではピッチ上昇により「自信がある」「元気がある」印象に影響を与え、ピッチ下降により「説得力がある」「丁寧である」印象に影響を与えると考え、4つの印象を評価対象とした。

4.2 実験用動画

上記の印象が重要になると予想される内容を取り入れた7種類の動画を用意した。会話内容は動画A、動画B、動画C、動画Dの4種類とし、動画A、動画B、動画Cについては2者の発言内容を入れ替えた2通りの会話動画をそれぞれ作成した。実験用動画では話し手と聞き手の2者が登場し、話し手が語尾ピッチ変化システムのユーザである。7種類の動画の概要は以下のようになっている。

動画A：旅行に行きたい場所

動画A-1は、話し手が旅行に行きたい場所の提案を行い、聞き手が応答する内容である。娯楽に関する話題であるため、元気であることが聞き手により良い印象を与えることにつながると予想される。動画A-2では話者を交代する。

動画B：分からないことについての相談

動画B-1は話し手が相談を持ち掛け、聞き手が応答するという内容である。相談をする際には丁寧な口調であることが聞き手により良い印象を与えることにつながると予想される。動画B-2では話者を交代する。

動画C：ディベート

動画C-1は話し手が議題を持ち掛け、聞き手が反論する内容のディベートを行う。発言内容に自信、説得力をもたせ、なおかつ丁寧な口調であることが聞き手により良い印象を与えることにつながると予想される。動画C-2では話者を交代する。

動画D：プレゼンテーション

話し手が資料を共有しながらプレゼンテーションを行う。プレゼンテーションでは発言に説得力・自信をもたせる必要があり、さらに丁寧な口調であることが聞き手により良い印象を与えることにつながると予想される。

動画A-Dの中で、「自信がある」「説得力がある」「元気がある」「丁寧である」の4つの印象のうち、話者が聞き手に与えたいと予想される印象をまとめたものを表2に示す。動画A-Dを1セットとして、1セット内で話し手と聞き手が同一人物になるようにしている。聞き手は固定せず、話し手を入れ替えて合計4セット（話し手4人×1セット）の撮影をおこなった。話し手の選定は20代学生男性3名と女性1名とした。動画A-Cでは図3(a)のように話し手と聞き手が表示される画面構成とした。動画Dでは図3(b)のように話し手と共有資料が表示される画面構成とした。

1セット内の7種類の動画それぞれについて、「加工無し」、「語尾ピッチ上昇」、「語尾ピッチ下降」の3種類のピッチ変化パターンを用意し、話し手1人あたり21種類の動画（3ピッチ変化パターン×7種類）を作成し、計84種

表 2: 動画内で話し手が聞き手に与えたいと考えられる印象

	自信	説得力	元気	丁寧
動画 A-1			○	
動画 A-2			○	
動画 B-1				○
動画 B-2				○
動画 C-1	○	○		○
動画 C-2	○	○		○
動画 D	○	○		○



(a) 動画 A-C

(b) 動画 D

図 3: 動画の画面構成

類（21種類×話し手4人）の動画を作成した。提案システムを再現した動画の作成手順として、まず話し手の音声と映像を0.9秒遅らせて聞き手に出力して行われた会話の音声と映像を録画し、提案手法を用いて加工した。

4.3 実験方法

実験開始時、被験者に話し手が語尾のピッチ補正をしていることを伝える。その後、被験者は椅子に座った状態で4.2節で述べた話し手1人分である21種類の動画をランダムな順序で視聴し、各動画を視聴するごとにアンケート調査を行った。アンケートでは被験者が動画のユーザに対して抱く印象について、4つの調査項目（Q1-Q4）を用意し、話し手に対して当てはまるものを回答させた。調査項目はQ1「自信があるように感じるか」、Q2「説得力があるように感じるか」、Q3「元気があるように感じるか」、Q4「丁寧であるように感じるか」の4つで、それぞれ（5: 感じる - 3: どちらでもない - 1: 感じない）の5段階のリッカート尺度で回答させた。被験者は21種類の動画を視聴後、自由記述式のアンケートを用意し、任意で回答を得た。話し手1人あたり被験者7名分の回答を得て、合計被験者28名分の回答（話し手4名×被験者7名）を得た。被験者は20代の学生21名である。

4.4 実験結果と考察

動画内容ごとのアンケート結果の平均値を図4に、ピッチ変化手法ごとのアンケート結果の平均値を図5に示す。エラーバーは標準偏差を示す。アンケート結果において、動画A-1と動画A-2、動画B-1と動画B-2、動画C-1と動画C-2とで平均値を取ったものをそれぞれ、動画A、動画B、動画Cとして、動画A-Dに対し動画内容、ピッチ変化に被験者内2要因分散分析を行った。Q1-Q4で動画内容

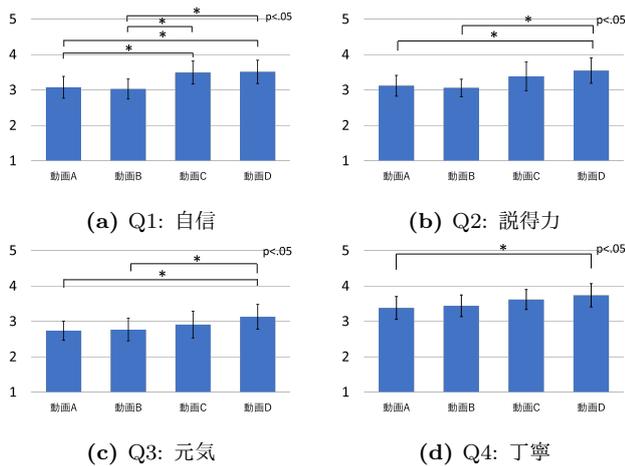


図 4: 動画内容ごとのアンケート結果

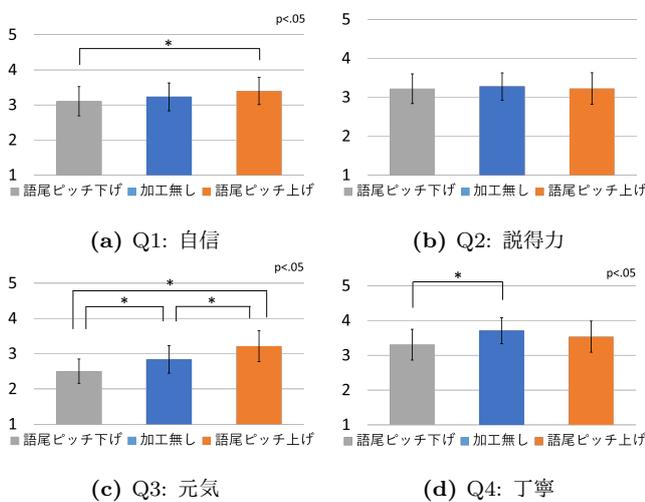


図 5: ピッチ変化手法ごとのアンケート結果

とピッチ変化の交互作用は有意差なしであった。以下では動画内容、ピッチ変化の主効果をみるため、Bonferroni法を用いて多重比較を行った結果を説明する。

Q1: 自信があるように感じるか

動画内容の主効果が有意 ($F(3, 28)=10.33, p<0.01$), ピッチ変化の主効果が有意であった ($F(2, 28)=6.96, p<0.01$). 動画内容におけるスコアの差は、図 4(a) のように、動画 A の平均値は動画 C 動画 D の平均値よりも有意に小さく、動画 B の平均値は動画 C 動画 D の平均値よりも有意に小さくなった。ピッチ変化手法におけるスコアの差は、図 5(a) のように、語尾ピッチ上げの平均値が語尾ピッチ下げの平均値より有意に大きかった。このため、提案システムを用いて自信を高める効果を求める場合、語尾ピッチ下げではなく語尾ピッチ上げを用いることが効果的であると考えられる。加工無しと語尾ピッチ上げとの間に有意な差は無いが、アンケート結果において加工無しに対して語尾ピッチ上げの平均値が高くなる傾向にあるため、提案システムを用いて自信を高める効果がある可能性がある。

Q2: 説得力があるように感じるか

動画内容の主効果が有意 ($F(3, 28)=5.11, p<0.01$), ピッチ変化の主効果が有意なしであった。動画内容におけるスコアの差は、図 4(b) のように、動画 A、動画 B の平均値が動画 D の平均値よりも小さかった。今回の結果からは、提案システムによる説得力を高める効果は確認できなかった。

Q3: 元気があるように感じるか

動画内容の主効果が有意 ($F(3, 28)=4.18, p<0.01$), ピッチ変化の主効果が有意であった ($F(2, 28)=21.81, p<0.01$). 動画内容におけるスコアの差は、図 4(c) のように、動画 A、動画 B の平均値が動画 D の平均値よりも小さかった。ピッチ変化手法におけるスコアの差は、図 5(c) のように、語尾ピッチ上げの平均値が加工無しの平均値より有意に大きく、語尾ピッチ上げの平均値が語尾ピッチ下げの平均値より有意に大きく、加工無しの平均値がピッチ下げの平均値より有意に大きかった。この結果から、提案システムによって元気な印象を与える、元気でない印象を与えること双方が可能だと考えられる。

Q4: 丁寧であるように感じるか

動画内容の主効果が有意 ($F(3, 28)=3.84, p<0.05$), ピッチ変化の主効果が有意であった ($F(2, 28)=5.48, p<0.01$). 動画内容におけるスコアの差は、図 4(d) のように、動画 A の平均値が動画 D の平均値よりも小さかった。ピッチ変化手法におけるスコアの差は、図 5(d) のように、語尾ピッチ下げの平均値が加工無しの平均値よりも有意に小さくなった。今回の結果からは、4.1 節で想定していた語尾ピッチ下げによって丁寧である印象を高める効果は確認できず、逆に語尾ピッチを下げると丁寧さが減る結果となった。

被験者から得られたアンケートから動画の種類によっては、ピッチ変化度合いが大きすぎる、音質が落ちているなど、ユーザに対して違和感を感じるという意見が得られた。オンライン会議において、話し手の音質が低いと話し手自身の評価が下がる [14] ことから、適切なピッチ変化度合いを検討することで結果が変わる可能性がある。

4.5 提案システムの実利用結果の考察

提案システムの実用性を調査するために、筆者自身が提案システムを長期間にわたって実際にオンライン会議で使用した。オンライン会議の内容は筆者が所属する研究室における進捗報告会である。会議は週に 1 回 2 時間程度の頻度で、およそ 1 年間システムを常に使用した状態で会議に参加し続けた。筆者自身の利用であるため客観的な評価は行えないが、システムを使用し続けた際に主観的に感じたこととして、音声と映像の遅延と、長期的な使用によるシステムへの意識の薄れが挙げられる。システム使用中は、自身の発話に対する相手の反応速度が遅れていたことによ

り、自身の音声が遅れて相手に届いていると感じることが多かった。また、システム使用開始当初は筆者自身がシステムに慣れていなかったため、システムを通して音声が出力されていることを意識していたが、使用を続ける中でシステムへの意識が薄れ、自身の音声加工されて出力されていると感じることは少なくなった。同会議に参加していたメンバに対し、ユーザである筆者に対して会議中に感じたことを自由記述式のアンケートを用いて調査したところ、「提案システムを使用し始めた当初は音声のディレイに違和感を感じていたが、回数を重ねていくと気にならなくなった」「システムありなしを比較していないので長期的な使用により印象が変化したかは不明であるが、提案システムの使用による成果報告に支障はなく、数ヶ月連続で提案システムを使用していると提案システムが使用されていることを意識しなくなった」という意見があった。これらから、提案システムの長期的な使用によりシステムへの慣れが生じ、提案システムによる語尾ピッチ変化はオンライン会議の進行に支障をきたすことなく使用できると考えられる。また、ピッチ変化への慣れが提案システムにより与える印象変化にどの程度影響するかについては調査ができていないため、今後長期的な実験の方法について検討する必要がある。

5. まとめ

本研究では、オンライン会議においてユーザが相手に与えたい印象を設定するとシステムがその設定に基づいてユーザ発話の語尾ピッチを自動で補正し、ユーザは普通に話すだけで意図する印象を聞き手に与えられるオンライン会議向け語尾ピッチ変換システムを提案した。オンライン会議において提案システムを利用した際に、聞き手が受ける印象の変化についての評価と、1年にわたる筆者のシステム利用経験の考察を行った。結果、提案システムと同じ条件で語尾ピッチを上昇させた場合、ユーザに対し元気である印象を抱く傾向が見られた。今後の課題として、提案システムの使用により聞き手に違和感を感じさせない適切なピッチ変化の度合いの調査を行う必要がある。また、提案システムによって操作可能な印象を明らかにすることで、聞き手により良い印象を与えるための音声補正を行うシステムの構築を目指す。

謝辞 本研究の一部は、JST CREST(JPMJCR18A3)、およびJST PRESTO(JPMJPR2138)の支援によるものである。ここに記して謝意を表す。

参考文献

- [1] A. Merabian: Communication Without Words, *Psychological Today*, Vol. 2, pp. 53–55 (Aug. 1968).
- [2] 市川 薫, 佐藤伸二: 対話理解に対する抑揚情報の役割, 情報処理学会研究報告音声言語情報処理, Vol. 57, pp. 51–58 (July 1994).
- [3] 郡 史郎: 間投助詞のイントネーションと間投助詞的イントネーション型の使い分けについて一, 言語文化研究, Vol. 42, pp. 61–84 (Mar. 2016).
- [4] 今井田亜弓: 若い日本人女性のピッチ変化に見る文化的規範の影響, 言語文化論集, Vol. 27, No. 2, pp. 13–26 (Mar. 2006).
- [5] T. Banziger and K. R. Scherer: the Role of Intonation in Emotional Expressions, *Speech Communication*, Vol. 46, pp. 252–267 (July 2005).
- [6] 鶴谷千春: 丁寧さを表現するために日本語母語話者が用いる韻律的特徴, 国立国語研究所論集, No. 11, pp. 167–180 (July 2016).
- [7] N. Dahlbäck, Q. Wang, C. Nass, and J. Alwin: Similarity Is More Important than Expertise: Accent Effects in Speech Interfaces, *Proc. of the SIGCHI Conference on Human Factors in Computing Systems (CHI '07)*, pp. 1553–1556 (Apr. 2007).
- [8] 常島杏菜, 吐師道子: 語尾上げ・助詞上げ発話の「STとしての専門職らしさ」の印象への影響, 音声言語医学, Vol. 58, No. 4, pp. 339–345 (Mar. 2017).
- [9] 亀岡弘和: 音声のイントネーションとアクセントを分析, 合成, 変換 (特集 未来への羅針盤としてのコミュニケーション科学), NTT 技術ジャーナル, Vol. 27, No. 9, pp. 10–12 (Sep. 2015).
- [10] T. Togawa, T. Otani, and K. Suzuki: Speech Enhancement Technology, Speech Rate Control Technology, *the Journal of the Institute of Electronics, Information and Communication Engineers*, Vol. 96, No. 11, pp. 874–881 (Nov. 2013).
- [11] E. Moulines and F. Charpentier: Pitch-Synchronous Waveform Processing Techniques for Text-to-Speech Synthesis Using Diphones, *Speech Communication*, Vol. 9, pp. 453–467 (Aug. 1990).
- [12] 小山哲春: 文末詞と文末イントネーション, 文法と音声, pp. 97–119 (May 1997).
- [13] 森山卓郎: 文の意味とイントネーション, 講座 日本語と日本語教育, Vol. 1, pp. 172–196 (Sep. 1989).
- [14] E. J. Newman and N. Schwarz: Good Sound, Good Research: How Audio Quality Influences Perceptions of the Researcher and Researcher, *Science Communication*, Vol. 40, No.2, pp. 246–257 (Apr. 2018).