聞き手によるパラ言語調整が対面対話に及ぼす影響の検討

田島　智宣[†]　小倉　加奈代[†]　西本　一志[†]

会話において，非言語情報は大きな役割を担っている．しかし，発言者が自身の音韻的特徴を客観的に評価する機会は少なく，また，聞き手の非言語行動が見過ごされる等，聞き手の会話への発信が機能しないことがある．本研究では聞き手によって話し手の音声を操作することで，聞き手の会話への働きかけを顕在化させ，それによる会話や関係への影響を調べる．

Effect of regulating paralanguage by listeners on conversations

TOMONORI TAJIMA[†]　KANAYO OGURA　KAZUSHI NISHIMOTO[†]

In conversation, paralinguistic information is important. However, speaker rarely estimates his own voice and its acoustic feature intentionally. Paralinguistic communication from listeners sometimes doesn't work because the speaker often misses the action of the listeners. In this study, we attempt to directly regulate the speaker's paralinguistic feature by the listeners based on the listeners' feelings without using non-verbal messages that are used in the usual communications. We investigate the effect of regulating the speaker's paralinguistic feature by the listeners in conversation.

1. はじめに

会話において音程や抑揚などの声色は発話に様々なニュアンスを加え，言葉で表現されている以上の感情や印象を相手に伝える．しかし，これらの要素は習慣の中で無意識に用いていることが多く，発言中に話し手自身が自己の発言の捉え方を客観的に分析することは難しい．そのため，話し手が思惑通りに，自分の音声を完全にコントロールできていない場合がある．例えば，地声が元々大きいために相手にとって威圧的だと感じられてしまうことや，常に早口で話してしまうため落ち着きがなく信頼感に欠ける人だと思われてしまうことがある．空気中を伝搬した声のみを，話し手が聞いているわけではないことが聞き手との差異を生み，客観視を難しくしている．このため　，アナウンサー等は，録音した自身の声を繰り返す聞く訓練を発声の上達のために取り入れている．

実際の会話では，自身の発言を直接評価するのではなく，話し手は主に自分の発言の影響を知る上で，聞き手の反応を参考にしながら，自分の発言を調整することが一般的である．ただし，その聞き手による反応は発言を妨げないよう，表情やうなずきといった仕草等，非言語的な反応で行われることが多い．しかし，聞き手がメッセージを込めた非言語行動を，話し手は見落としてしまう恐れがある．また，見ていてもその行動に込められた聞き手からの情報を読み取らなかったり，読み取ることができなかったりすることがある．そのために俗に言う「空気が読める・読めない」などという差異が生まれる．話し手が聞き手からの情報をくみ取らず，話し手からの一方的な発信に留まる会話は，聞き手にとって非常に苦痛である．

このような状況を解決するべく，本研究では，「聞き手の非言語行動を話し手が読み取り，解釈した上で，話し手自身が声に反映させる」という間接的で煩雑なやり取りを簡易化し，「聞き手が，本来非言語的に伝えていたメッセージを，直接話し手の音声に反映させる」というやりとりに変更し，それによって会話にどのような影響が現れるのかを調べる．具体的には会話参加者全員に変声機を着用してもらい，その操作を聞き手のみが行えるようにするシステム、Marionette Voice を構築し、これを用いて実験と評価を行う。

以下，2 章では本研究にかかわる関連研究を概観する．3 章では構築した Marionette Voice について解説し，4 章にて実験の内容と方法，結果を述べ，5 章にて実験結果に基づき考察する．6 章にてまとめと今後の課題について述べる．

† 北陸先端技術大学院大学 知識科学研究科
School of Knowledge Science
Japan Advanced Institute of Science and Technology

2. 関連研究

2.1 会話と音韻的特徴

Apple,Streeter & Krauss(1979)[1]では声のピッチと発話速度に注目し，説得会話の中で発話者に抱く印象とこれらの要素の関係について調べている．声のピッチが高いほど，誠実でなく，力強さがなく，神経質に感じられ，説得力がなかった．また，遅く話すほど誠実でなく，流暢さにも欠け，説得力もなく，消極的に感じられた．

声の高さ等の音韻的特徴と，それと結びつく感情に関する研究は人工音声や音声認識の分野において特に盛んである．Johan(2010)[2]では，感情とそれを表現する音韻的特徴の関係を明らかにし，感情を検知するシステムへの応用の可能性を述べている．

これらの研究では，音声と印象や感情との関係性を調べているが，予め，ピッチや速度を調整した音声を聞いてもらい，その音声に対してどういう印象や感情を抱くのか調べる実験を行っている．本研究では，聞き手の持つ印象や感情によって音声を変更した際に，同様のことが言えるのかを確かめる．

2.2 会話と非言語行動

Goodwin(1981)[3]は，会話における言語行動と身体行動の関係についてビデオ分析を行っている．レシピについて尋ねるという短い日常場面の中で言語的発話，会話の構造，視線や身体状態といった様々な要素がどのように組み合わさっているのかについて調べている．Maynard(1993)[4]は，うなずきの頭の動き方をパターン化し，そこに込められた意味に関する解析を行っている．また，会話における頭の動きについて日米の差も比べており，機能としては類似していることを挙げるとともに，その頻度においては大きな違いがあることを述べ，このような非言語行動も使用する言語や文化によって差異があることを指摘している．

聞き手による話し手の音声を変更する行為は，上記のような聞き手の行動が非常に顕在化され直接的になったものであると考えられ，これらと同列の役割を持つか比較し，変声行為を通じた新たなコミュニケーションの可能性について検討する．

3. 実験システム

構築した Marionette Voice は，聞き手が話し手の声を変声しながら会話を行うシステムである．図 1 に，Marionette Voice の全体図を示す．Marionette Voice は変声・パラメータ操作・通信を行うクライアントと，

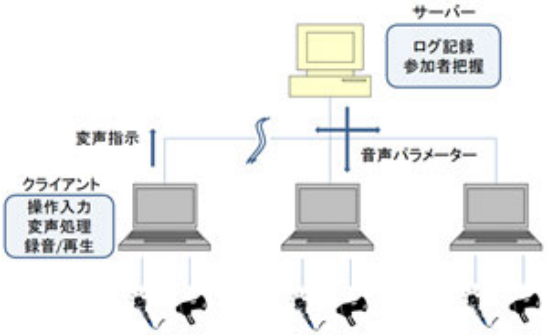

図 1: システム全体図

通信・記録を行うサーバで構成されている．

クライアント上にて変声対象者を選択し，その対象者の発声のピッチと音量を操作することができる．ただし，自分自身を変声対象者とすることはできない．また，誰が変声操作を行ったのか各ユーザには通知しないため，変声操作は操作した本人にしかわからない．

変声操作によって変化した各パラメータはサーバを介して各クライアントに伝えられ，ユーザ別に保存される．受け取ったパラメータ変化の対象者がそのクライアントを操作しているユーザ自身であればそのパラメータに応じて，変声がなされる．受け取ったパラメータ変化の対象者が，そのクライアントを操作しているユーザ以外の場合は，現在のパラメータ表示に反映される．変声操作は最後の変更のみが反映され，常に最新の更新で上書きされる．

録音や再生はオープンソースで開発されているオーディオ I/O ライブラリ「Port Audio」[5]を用いた．また，変声のためのフーリエ変換にはオープンソースで開発されている FFT ライブラリ「FFTW」[6]を用いた．プログラミング言語に関しては，主に C++を用い，UI に関する部分では C++/CLI を用いた．

図 2 に，クライアントのインタフェースを示す．ユーザは，サーバの IP アドレスと自身の名前を所定のテキストボックスに記入した後，開始ボタンを押して会話を開始する．開始後は変声対象ドロップタブから声を変えたい相手を選ぶ(自分自身以外を選択可能)．スライダーを使って，50%から 150%まで対象者の声のピッチと音量を変更できる．変更は送信ボタンを押すことでサーバを経由して全員に送信され反映される．ピッチ変化スライダーと音量変化スライダーそれぞれの下のバーは，現在選択中の対象者に関するピッチと音量の設定状態を表している．

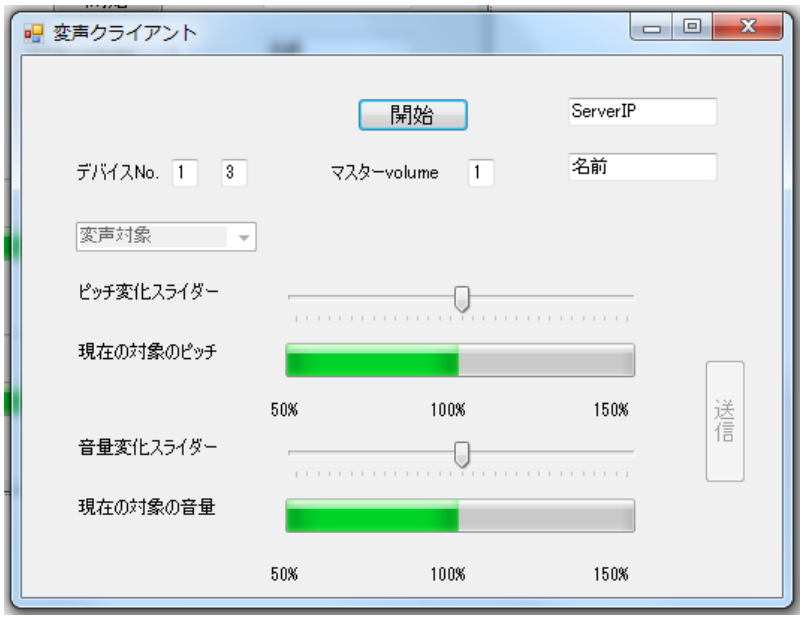

図 4:クライアントのインタフェース

4. 実験

4.1 概要

聞き手の行ったピッチと音量の変更に傾向はみられるのか，また話し手はそれを受けてどのように発話を続行していくのかを評価するために実験を行った．実験では，ノートパソコン・スピーカー・マイクのセットを人数分用意して，被験者に会話をしてもらった．ノートパソコン上にてクライアントを起動し，録音，変声，再生，通信，操作を管理した．スピーカーからは変声された声のみが出力される．地声を会話参加者に聞かせないために喉の振動で音声を取得するスロートマイク(図 3)を使用し，さらにマスクを装着した（図 4).

4.2 手順

被験者は 3 人であり，全員 24 歳の男性である．被験者はもともと友人関係にあったため，会話のテーマは決めずとも会話が発生すると判断し，特にテーマを設けずに雑談をしてもらった．説明とシステムへの慣れのためにまず 15 分ほど自由に操作してもらい，その後実験を約 40 分間行った．実験風景を図 5 に示す.

4.3 結果

表 1 に，被験者が送信したパラメータとその操作回数を示す．ピッチ・音量に関して，元の音声と比べて，高い状態か，低い状態か，元の音声と同じか，の 3 つのどの状態に設定して全員に向かって送信したかの回数を示している．ピッチに関しては，元音声より高い状態の送信が 53 回，元音声と変わりない状態の送信が 22 回，元音声より低い状態の送信は 27 回と，ピッチを上げる状態の送信が半分を占めていた．音量に関しては，音量を上げた状態の送信が 55 回，元音声と変わりない無い状態の送信が 40 回，音量を下げた状態の送信が 7 回であった．音量に関しては，元と同じ

図 2:スロートマイク

図 3:マスク

図 5:実験風景

表 1　送信の内訳

	音量				
		上昇	元音声	下降	計
ピッチ	上昇	25	23	5	53
	元音声	13	7	2	22
	下降	17	10	0	27
	計	55	40	7	102

か高い状態の送信が 9 割以上を占めていた．

また，話し手の音声の状態の変遷について図 6～8 に示す．元の音声より高い状態に設定されていた時間は 3 人平均で 40 分中 30 分 30 秒と 76%にもおよび，元の音声から変更がない場合も含めると 35 分 23 秒と 88%となり，会話全体において 9 割以上もの時間の間元以上のピッチであった．音量に関しても 9 割以上が元以上の大きさであった．

被験者たちが変声した時の理由をアンケートによって調べたところ，3 名とも聞きづらい時と答えた．また，うち 2 名は他にも相手の声を面白くするためや暇なときに遊びとしても利用したという回答を得た．

変声内容に関しては，2 名が低い声は威圧感があり，高いと親しみやすいと答えた．残りの 1 名はピッチの変化対しては特に傾向は感じないと答えた．音量に関しては聞こえやすさのみが変化し，他の傾向は感じられないとの回答を得た．

話し手として自身の声が変声されたことに関して尋ねてみたところ，特に意識は向けられておらず，変声を受けて意識や行動が変わるようなことは無かった．

なお，変声に対する興味から，とりあえず声を出して確かめたり，お互いの変わった声を楽しんだりするために意味や脈絡のない声を発する等の行為が実験の最初の頃や会話が滞った際に集中して見られた．

5.　考察

変声内容がピッチも音量も上げる方向に偏っているのは，操作目的が聞きづらい時に聞きやすくするためだったと考えられる．低い声は怖いというネガティブなイメージがあった半面，高い声に対しては中立からややポジティブな意見がアンケートから見て取れた．変声の理由は，何か感情や印象を左右するのではなく，聞き手が心地よい，会話をする上で最も良い声を探すために操作していた．ただし，これは話者の生の声を遮って，スピーカーを通して会話するシステム自体が既に普段の会話とはあまりに違うため，聞き取りづらく，そのために聞き取りやすくすることが前提になってしまった可能性は否めない．

また，今回の実験ではお互いの立場に差がなく，仲の良い間柄であったため会話に積極的であったが，反目し合っている間柄や，上下関係のある間柄など，会話に対して否定的だったり，消極的だったりした場合には，聞きやすくするという目的と操作傾向も変わる可能性がある．

変声行為自体に面白みを感じ，それが目的となってしまって，会話を阻害するような変声も見られた．ピッチを変化させても具体的にどのような声になるのか想像がつかないため，しばらくは興味本位で試しながら操作するためである．だが，聞き取りづらい低い声への変声は操作回数こそ 25%程度あったものの，実際の時間にすると 10%未満になっていることから分かるように，聞き取りやすい高い声へと短時間で変更されていることがわかる．興味本位の行動は長続きせず，参加者の会話への参加意欲がこれらを打ち消す方向に働いている可能性がある．よって，このシステムに慣れてきて，ピッチの変更に意味づけがなされてくれば，次第にこのような行為は減っていくと考えている．

変声を受けて，話し手は特に行動を変えるようなことは特には観察されず，変声行為を通じて話し手と聞き手の間にどのような新規なコミュニケーションが形成されるかに関しては依然不明である．今回のような短時間の使用では，まだ変声に慣れておらず，聞こえやすくする以上の意味形成がなされていなかったためだと思われる．特に話者に対してピッチの変更を求めることは通常会話ではあまりなされないため，この意味形成はさらなる時間が必要となると思われる．今後さらにシステムを継続的に使用することによって，変声行為への新たな意味づけが創発することが期待される．

6. まとめ

聞き手が話し手の音声のピッチと音量を直接的に変化させたとき，その操作がどのように行われるのか，その結果として会話がどのような影響を受けるのかを調査する実験を行った．その結果，聞き手は聞こえてくる音声ができるだけ聞きやすいように調整する傾向があることが分かり，また，それはピッチと音量を上昇させる操作であった．しかしながら，現時点ではまだこのような聞き手による話し手の音声の直接的な変更に対する意味づけがなされておらず，聞きやすさの調整以上の意図や，これによる話し手と聞き手の間での意思疏通までは確認できなかった．

本研究が最終的に目指す，話し手からの一方的な発信に留まる会話を避け，聞き手の意向をより直接に反映できる会話を実現可能かどうかを明らかにするためには，さらに長期にわたる実験を行う必要がある．また，現在は最新の変更しか反映しない状態で行っているが，これが本当に正しいのか確かめるべく，その場の聞き手の全員の意思を同時に反映する場合（たとえば平均値を用いる場合）の実験を実施し，今回の結果と比較したい．また，今回は参加者の立場が全員ほぼ対等であったため，会話に特定の方向性は特になかった．明確な会話上の役割の違いや，実際の人間関係上での立場の差などをこの会話に持ち込んだ際に，システムの利用方法に差が出るのか，検証したいと考えている．

謝辞　本研究の一部は，平成 21 年度（財）栢森情報科学振興財団の研究助成を受けて実施された．ここに謝意を表する．

参　考　文　献

1) William Apple, Lynn A. Streeter, Robert M. Kraussa Effects of pitch and speech rate on personal attributions　Journal of Personality and Social Psychology　Volume 37, Issue 5, May 1979, Pages 715-727
2) Sundberg, J. Interdependencies among voice source parameters in emotional speech Affective Computing, IEEE Transactions on July-Sept. 2011 2 Issue:3 162 - 174
3) Goodwin, C. (1981). Conversational Organization: Interaction between Speakers and Hearers. New York: Academic Press.
4) 泉子・K・メイナード　「会話分析」くろしお出版 pp152-179
5) PortAudio　http://www.portaudio.com/
6) FFTW　http://www.fftw.org/

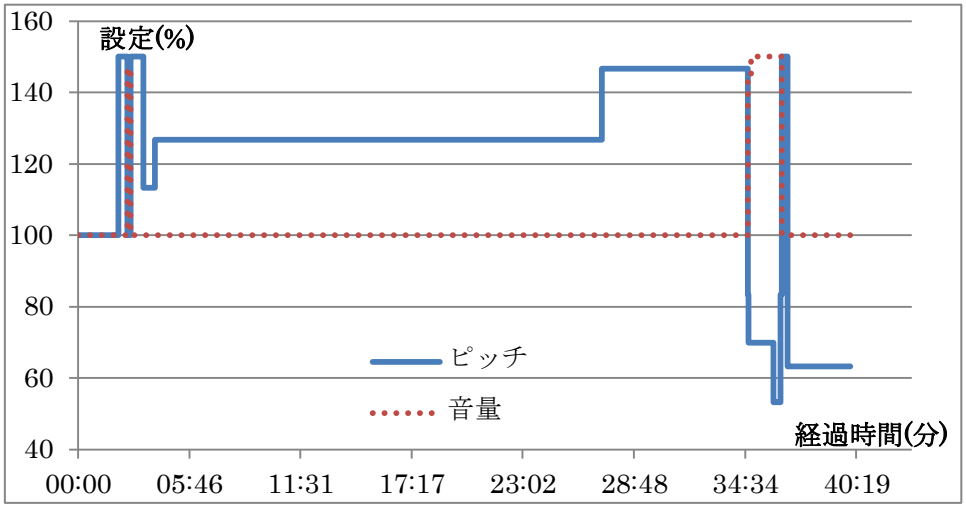

図 6 被験者 A の時間変移

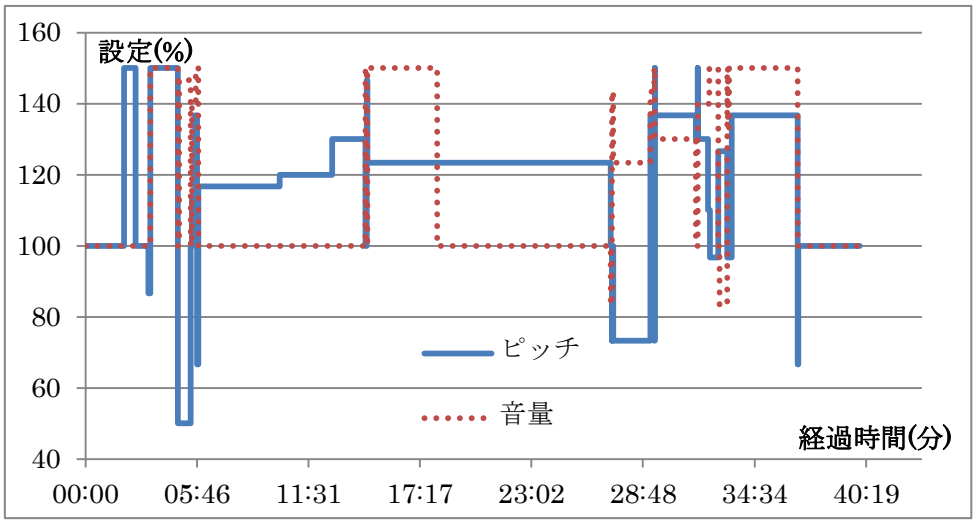

図 7 被験者 B の時間推移

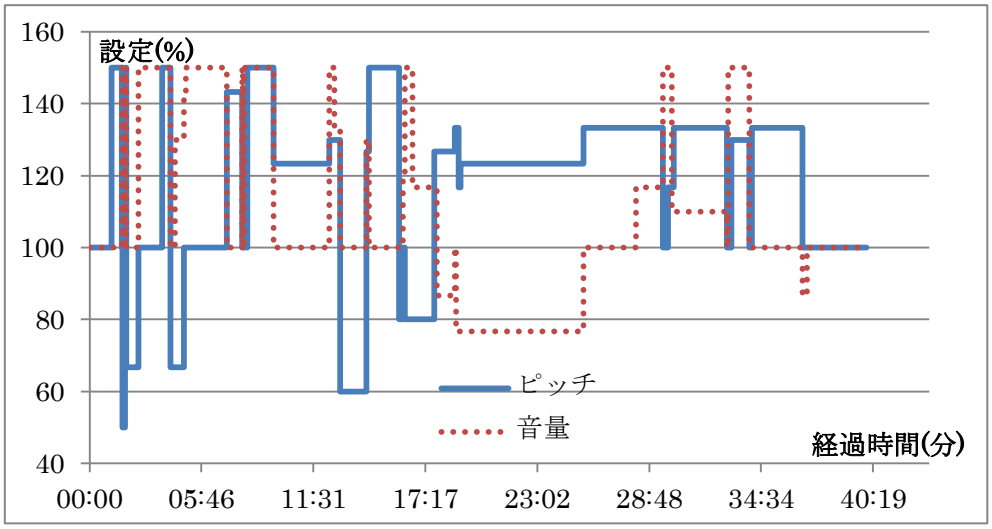

図 8 被験者 C の時間推移