法政大学学術機関リポジトリ
HOSEI UNIVERSITY REPOSITORY

PDF issue: 2025-07-12

発話速度に変化があったと感じながらも不自然ではない持続時間の範囲とその範囲内で変化する評価とは ： 多段階に持続時間の異なる合成音声を用いた音声知覚と推論の検討

橋本, 和奈実

(出版者 / Publisher)
法政大学大学院
(雑誌名 / Journal or Publication Title)
大学院紀要 = Bulletin of graduate studies
(巻 / Volume)
89
(開始ページ / Start Page)
35
(終了ページ / End Page)
45
(発行年 / Year)
2022-10-31
(URL)
https://doi.org/10.15002/00026027

発話速度に変化があったと感じながらも不自然ではない持続時間の範囲とその範囲内で変化する評価とは
――多段階に持続時間の異なる合成音声を用いた音声知覚と推論の検討――

Exploring a range of duration that feels like there has been a change in speaking rate but is not unnatural and changes in evaluation within this range: Investigation of speech perception and inference using synthetic speech with multi-step duration

人文科学研究科　心理学専攻
博士後期課程 3 年　橋本　和奈実

これまで持続時間を変化させた合成音声を用いた多くの実験研究は，実験デザインの関係上，少数の水準を用い，音声知覚や推論の検討を行っている。その一方で，人が発話速度に変化があったと感じながらも不自然ではない持続時間の範囲については，十分に検討されていない。本研究の目的は，発話速度に変化があったと感じながらも不自然ではない持続時間の範囲を知り，その持続時間の範囲の中で，音声知覚と推論の議論を行うことである。目的達成のために，「こんにちは，初めまして，よろしくお願いします。」と発話している音声の持続時間を段階的に 19 水準の変更を加えた音声を作成した。そして作成した音声を参加者に聞かせ，音声の速さ感・高さ感・自然性の知覚と，話者の年齢と悠長さの推論を尋ねた。その結果，10.23 モーラ/s から 5.28 モーラ/s の間では，人が発話速度の変化があったと感じながらも，音声の自然性が保たれる音声の合成範囲となることが明らかになった。また，その持続時間の範囲の中で，悠長さの推論が有意に変化することが認められた。本研究で示された持続時間の指標を基にすれば，本研究では扱うことのできなかった音声知覚や推論に対しても，引き続き同様の検討をすることができる。

キーワード：自然性，速さ感，高さ感，年齢推定

Many previous studies that used synthetic speech stimuli have artificially manipulated duration to investigate speech perception and speaker inference. These studies have often run experiments with a small number of duration conditions, e.g. slow, normal, fast, because of other constraints. However, the choice of duration conditions is often arbitrary; range of duration that feels like there has been a change in speaking rate while still sounding natural has not been directly investigated. The purpose of the present study is to determine the range of duration in synthesized speech that is wide enough yet natural, and to explore speech perception and speaker inference within such a range. To achieve this goal, duration of recordings of a typical first-time greeting in Japanese was manipulated in 19 steps. Participants listened to the stimuli and evaluated the perceived speed, pitch, and naturalness of the utterance, along with inference of age and voice impression. Results revealed that a range of duration that feels like there has been a change in speaking rate but is not unnatural was 63.51% - 128.45% of the original rate (10.23 – 5.28 moras per second). Within this range of duration, inference of voice impression significantly changed. These results provide an empirical basis for selecting appropriate ranges of duration variation for investigating aspects of speech perception and speaker inference.

key words: naturalness, perceived speed, perceived pitch, age estimation

Correspondence concerning this article should be sent to: Kanami Hashimoto, Major in Psychology, Graduate School of Humanities, Hosei University, Ichigayatamachi, Shinjuku-ku, Tokyo 162-0843, Japan. (E-mail: qwe783jd@gmail.com)

問題と目的

人は他者とコミュニケーションをとるとき，音声を用いなければならない場面が多々ある。特に対面以外でのコミュニケーションを求められてきた昨今，例えば，オンラインで音声通話をする機会も世界的に増えている。

大坊（1998）によると，音声は，言語的な側面とパラ言語（非言語）的な側面を持っている。特にパラ言語には，多くの要素が存在する。例えば，平均基本周波数や持続時間，声の大きさ，間の置き方が，パラ言語に含まれる。

聴者は話者のパラ言語を基に，話者に対する情報をまとめようとする。例えば，情動の推測や印象の形成を行う（e.g Allport & Cantril, 1934 ; 藤原， 1986 ; Scherer, Banse, Wallbott, & Goldbeck, 1991）。

合成音声を用いた研究

技術の発展に伴い，音声合成ソフトが開発されることで，音声の中の特定のパラ言語を限定的に操作することが可能となった（e.g. Praat Ver.5.4.15（Boersma & Weenink, 2015））。音声信号の情報を操作した音声を合成音声と言い，このような合成音声を用いた研究も行われ始めている（e.g. 笠原・厳島，2007 ; 内田，2002）。

合成音声を用いた研究を，いくつか紹介する。内田（2000）は，平均基本周波数や持続時間を操作した合成音声を用い，音声の知覚や性格特性の印象を測定する実験を行った。その結果，平均基本周波数については，平均基本周波数が高くなると聴者が感じる高さ感が増すが，それだけではなく，速さ感も増すことを明らかにした。持続時間については，持続時間が長くなると聴者が感じる声の速さ感が遅くなることに加え，基本周波数が変化していないにもかかわらず，声の高さ感も低くなることが明らかになった。内田（2005）は持続時間の中でも，特に，実際の発話時間に着目し，回帰式の当てはめを行った。その結果，自然性は，中央付近が頂点の逆 U 字型の 2 次式に最も当てはまりが良かった。また，速さ感（内田（2005）では発話速度感としているが，本研究内では速さ感と表現する）は，右肩下がりで U 字型の 2 次式に最も当てはまりが良かった。橋本・古屋（2019）は，平均基本周波数と持続時間を操作した合成音声を用い，音声や性格特性の印象を測定する実験を行った。その結果，平均基本周波数と持続時間によって，音声の印象や性格特性の印象の推測が異なることを明らかにした。

先述した合成音声を用いた研究は，実験デザインの関係上，少数の水準を用いた検討となっていた。そのため，パラ言語の操作範囲が狭すぎてパラ言語が変化したと感じにくかったり，反対に，広すぎて普段の会話ではあまり聞くことのない不自然なパラ言語となっていたりした可能性が否定できない。また，パラ言語の変化による話者への推測の細かな違いを調べるには至っていない。これらの問題を解決するための手段の一つとして，人がパラ言語の変化があったと感じながらも，普段耳にするパラ言語から逸脱しない音声の合成範囲を知ることが有効となるだろう。その上で，操作していないパラ言語の知覚が変化する現象についてや，話者への推測についての議論を行うことが望ましいだろう。

この観点に立ち，橋本（2020）は，広範囲かつ細かく，19 段階に平均基本周波数を変化させた合成音声を用いた実験を行った。声の高さ感に変化があったと感じながらも不自然ではない平均基本周波数の範囲を知るために，高さ感・自然性を測定した。パラ言語知覚の議論のために，速さ感を測定した。推測の議論のために，話者の年齢推定・話者の気分の推測を測定した。結果は，1—3 次までの回帰曲線との当てはまりの良さを調べる回帰分析によって取りまとめた。回帰分析は，平均基本周波数を対数変換した値と，それぞれの得点に対して行った。橋本（2020）では，得られた回帰式を利用することで，声の高さ感に変化があったと感じながらも不自然ではない平均基本周波数の範囲を知ることができた。また，その平均基本周波数の範囲の，速さ感・話者の年齢推定・話者の気分の推測の変化に対して，一部の議論を行うことができた。

本研究の目的

本研究は，平均基本周波数に着目した橋本（2020）に続き，基本周波数以外のパラ言語に着目した同様の研究を行うことを目指す。これまでパラ言語と知覚・推測に関する研究の多くは，平均基本周波数や持続時間に注目してきた（e.g. Apple, Streeter, & Krauss, 1979；丸島，2008）。そこで本研究では，基本周波数以外のパラ言語として，持続時間に着目する。実験条件と刺激は橋本（2020）を参考にし，19 段階に持続時間を変化させた合成音声を用いる。この 19 段階の条件の持続時間は，具体的には，100%条件と，100×(1+X/100)%あるいは 100÷(1+X/100)%と対数倍になる（X には 8×（1—9）が当てはまる）ように時間伸縮率を変化させた 18 条件から構成する。測定する内容は，速さ感・高さ感・自然性・話者の年齢推定と，新たに加える，悠長さとする。悠長さは，音声に対する印象や音声から推測できる話者の発言中の状態に関する印象をまとめた音声状態印象の中の 1 因子であり，持続時間との関連が，橋本・古屋（2019）によって示されている。話者の気分の推測については，橋本（2020）で十分な結果を得ることができなかった。そこで，十分な結果を得られない可能性を本研究において回避するために，話者の気分の推測を測定せず，その代わりとして，本研究では悠長さを加えた。また，橋本（2020）の課題として，回帰分析のみでは得られたデータを生かしきれず，音声知覚と推測の議論を十分に行えなかったことも挙げることができる。そこで本研究では，回帰分析に加えて，*t* 検定を行う。具体的には，まず，音声の自然性が保たれる音声の持続時間を，自然性の回帰式を用いて算出する。持続時間の条件の性質と，内田（2005）から，自然性の回帰曲線は中央付近が頂点の逆 U 字型となると予想する。そのため，音声の自然性が保たれる限界の音声の持続時間は二つ得られるだろう。その後，算出した持続時間の範囲内かつ二つ（最小および最大）の持続時間に最も近い持続時間となる 2 条件を選択し，高さ感の知覚や，話者の年齢推定・悠長さの推測の指標に対して，*t* 検定を行う。*t* 検定を行うことにより，発話速度に変化があったと感じながらも不自然ではない持続時間の範囲中での，高さ感の知覚や，話者の年齢推定・悠長さの推測に対しての議論が可能となる。

本研究の目的は，発話速度に変化があったと感じながらも不自然ではない持続時間の範囲を知り，その持続時間の範囲の中でパラ言語知覚と推測の議論を行うことである。発話速度に変化があったと感じながらも不自然ではない持続時間の範囲を知るために，速さ感・自然性を測定する。パラ言語知覚の議論のために，高さ感を測定する。推測の議論のために，話者の年齢推定・悠長さを測定する。目的を達成するために，3 次までの回帰式の当てはまりの良さを調べる回帰分析と，*t* 検定を行う。回帰分析と *t* 検定の結果を基に，発話速度に変化があったと感じながらも不自然ではない持続時間の範囲を示すと共に，その持続時間の範囲の中で高さ感の知覚や話者の年齢推定・悠長さの推測についての検討を行う。

方　法

倫理審査

本研究は，法政大学文学部心理学科・心理学専攻倫理委員会の承認を得て行った（承認番号 15-0118）。

実験計画

本研究の実験計画は，持続時間（58.14%（1.95 s），60.98%（2.05 s），64.10%（2.15 s），67.57%（2.27 s），71.43%（2.40 s），75.76%（2.55 s），80.65%（2.71 s），86.21%（2.90 s），92.59%（3.11 s），100.00%（3.36 s），108.00%（3.63 s），116.00%（3.90 s），124.00%（4.17 s），132.00%（4.44 s），140.00%（4.70 s），148.00%（4.97 s），156.00%（5.24 s），164.00%（5.51 s），172.00%（5.78 s））による，1 要因 19 水準参加者内計画であった。従属変数は，速さ感・高さ感・自然性・話者の年齢推定・悠長さであった。

参加者

参加者は，母語が日本語で聴覚に障害のない 21 名（男性 10 名，女性 11 名，M = 19.05，SD = 0.97）であった。21 名の参加者は，実験者から実験内容，実験参加の同意，データや個人情報の管理に関する説明を受けた上で，自主的に参加した。なお，音声に関する専門知識の提供や事前訓練は行わなかった。

実験材料

本研究には，PC（ARROWA Tab Q555/K64），ヘッドホン（SENNHEISER HD270 Control），Super Lab（Ver.4.5）を用いた。

本試行で用いた実験音声

原音声は，橋本（2020）と同様の音声であった。橋本（2020）では，母語が日本語で聴覚と言語に障害のない 15 名（男性 7 名，女性 8 名，$M = 22.13$，$SD = 0.92$）の音声を収録した。収録した台詞は 28 個あった。1 名の話者につき 1 から 4 回，休憩を挟みながら収録を行った。収録中に発話困難となった 1 名を除いた 14 名（男女各 7 名, $M = 22.07$, $SD = 0.92$）の音声の中で, 第一印象の重要性が指摘されていることを考慮し（Asch, 1946），初対面挨拶場面の台詞である「こんにちは，初めまして，よろしくお願いします。」と発話している音声を用いた。これらの音声の平均基本周波数と持続時間を，Praat Ver.5.4.15（Boersma & Weenink, 2015）を用いて測定した。平均基本周波数については，話者の性別ごとに平均値を算出した。持続時間については，話者全員の，発話全体の持続時間の平均値を算出した。なお，事前に，話者の性別を要因とした t 検定を行い，平均基本周波数には有意な差が認められ，発話全体の時間・実際に発話している時間・ポーズの時間に有意な差が認められないことを確認した。話者の性別ごとに算出した平均基本周波数の平均値と，話者全員の発話全体の持続時間の平均値を基に，収録した 14 名の音声の中で，最も平均基本周波数と発話全体の持続時間の平均値に近い音声特徴を持つ，男性話者と女性話者の 2 名（男女各 1 名，23 歳）の音声を選定した。

橋本（2020）の実験条件の設定に倣い，持続時間を算出し，本試行で用いた実験音声を作成した。本研究での持続時間 100%条件は，先ほど算出した，話者全員の，発話全体の持続時間の平均値である 3.36 s とした。残りの 18 条件の持続時間は，橋本（2020）の実験条件の設定に倣い，100%条件に対して，100×(1+X/100)あるいは 100÷(1+X/100)と対数倍になる（X には 8×（1—9）が当てはまる）時間伸縮率と決定し，この値をそれぞれ算出した。なお，18 条件の持続時間は実験計画で示したため，ここでの列挙は割愛する。平均基本周波数は，橋本(2020)で平均的な声の高さだと評価された 100％条件の平均基本周波数(男性は 137.83Hz, 女性は 263.43Hz)に統一した。100%条件の持続時間および算出した 18 条件の持続時間と，橋本（2020）に基づいて決定した平均基本周波数となるように，選定した男女各 1 名の音声の持続時間と平均基本周波数を，Praat Ver.5.4.15（Boersma & Weenink, 2015）の ManipulationEditor を用いて操作した。こうして作成した 38 音声（19 条件×話者 2 名分）を，実験音声とした。なお，Praat Ver. 5.4.15（Boersma & Weenink, 2015）の限界から，平均基本周波数については 0.49 Hz 以内の誤差が生じた。

持続時間の異なる音声のみでは参加者に実験の意図を読み取られてしまう可能性があるため，平均基本周波数の異なるダミー音声を作成した。ダミー音声の平均基本周波数の条件は，高音条件と低音条件の 2 条件であった。具体的には，男性話者の場合は高音条件が 192.49 Hz で低音条件が 99.32 Hz であり，女性話者の場合は高音条件が 367.51 Hz で低音条件が 188.23 Hz であった。これらの平均基本周波数は，橋本（2020）の結果から，自然性を保ちながら平均基本周波数の変化があったと感じながらも，音声の自然性が保たれる範囲と言える平均基本周波数であった。ダミー音声の持続時間は，全て，100%条件である 3.36 s とした。ダミー音声の平均基本周波数と持続時間も，Praat Ver.5.4.15（Boersma & Weenink, 2015）の ManipulationEditor を用いて操作した。作成したダミー音声は，4 音声（平均基本周波数 2 条件×話者 2 名分）であった。Praat Ver.5.4.15（Boersma & Weenink, 2015）の限界から，0.55 Hz 以内の誤差が生じた。

練習試行で用いた実験音声

原音声は，本試行で用いた実験音声と同様の話者 14 名の音声の中で，橋本（2020）で収録した 28 個の台詞の中の一つである「すぐに返しますので，ノートを貸してください。」と発話している音声であった。本試行で用いた実験音声と同様に，14 音声の平均基本周波数と持続時間を，Praat Ver.5.4.15（Boersma & Weenink, 2015）を用いて測定し，話者の性別ごとに平均基本周波数の平均値と，話者全員の発話全体の持続時間の平均値を算出した。算出した結果，男性の平均基本周波数の平均値は 139.42 Hz，女性の平均基本周波数の平均値は 246.17

Hz，話者全員の発話全体の持続時間の平均値は 2.87 s であった。練習試行で用いた実験音声は，14 音声の中で平均基本周波数と発話全体の続時間の平均値が総合的に最も離れていた，本試行で用いた音声の話者 2 名とは異なる，話者 2 名（男女各 1 名，$M = 22.50$，$SD = 0.71$）の音声とした。

平均基本周波数と発話全体の続時間の平均値を基準（100%）として，ダミー音声の平均基本周波数の高音条件（男性話者は 195.18 Hz，女性話者は 344.64 Hz）と低音条件（男性話者は 99.58 Hz，女性話者は 175.84 Hz）と，本試行で用いる実験音声の持続時間の条件の中で，最も極端な条件である 58.14%条件（1.67 s）と 172.00%条件（4.93 s）の 2 条件を組み合わせて，4 音声を作成した。4 音声の条件の内訳は，（a）男性話者・平均基本周波数が低音条件（99.58 Hz）・持続時間が 58.14%条件（1.67 s），（b）男性話者・平均基本周波数が高音条件（195.18 Hz）・持続時間が 172.00%条件（4.93 s），（c）女性話者・平均基本周波数が高音条件（344.64 Hz）・持続時間が 58.14%条件（1.67 s），（d）女性話者・平均基本周波数が低音条件（175.84 Hz）・持続時間が 172.00%条件（4.93 s）であった。あくまでも練習試行であるため，考え得る全ての組み合わせである 8 音声（話者 2 名×持続時間 2 条件×平均基本周波数 2 条件）とはしなかった。Praat Ver. 5.4.15（Boersma & Weenink, 2015）の限界から，4.97 Hz 以内の誤差が生じた。

質問項目

本研究では，速さ感・高さ感・自然性・話者の年齢推定・悠長さを測定した。速さ感を測定するために「遅い–速い」を，高さ感を測定するために「低い–高い」を，自然性を測定するために「不自然な–自然な」を，各項目の左側（e.g. 遅い，低い）を 1，右側（e.g. 速い，高い）を 7 とした，7 件法で用いた。話者の年齢推定は，質問紙に整数を記入することで測定した。悠長さは，橋本・古屋（2019）で特に持続時間との関連があった SD 尺度 3 項目（「ゆっくり話す–早口な」「せっかちな–ゆったりとしている」「ぞんざいな–丁寧な」）を，各項目の左側（e.g. ゆっくり話す，せっかちな）を 1，右側（e.g. 早口な，ゆったりとしている）を 7 とした，7 件法で用いた。

手続き

実験は防音室で個別に行った。まず，実験内容，実験参加の同意，データや個人情報の管理に関する教示を行い，同意書にサインを求めた。同時に，参加者に母語が日本語であること，聴覚に障害がないことを確認した。その後，性別と年齢を質問紙によって尋ねた。これらの事前説明および確認の後，実験を開始した。

実験は，音声の知覚や印象（速さ感・高さ感・自然性・悠長さ）を尋ねるセクションと，話者の年齢推定を尋ねるセクションに分かれていた。両セクションとも，参加者に音声を 1 音声ずつ聞かせ，質問項目への回答を求めた。また，両セクションとも，参加者への音声の呈示は 1 度だけであった。音声呈示の前には，1000 Hz の純音を 500 ms，両セクションとも呈示した。加えて，両セクションとも，男性話者の音声と女性話者の音声とを前後半で分けて呈示し，話者の性別が変わる際には PC 画面上で参加者に知らせた。音声の流れる順番はラテン方格を用いてコントロールし，両セクションとも同じ順番で流した。両セクションとも，本試行に入る前に練習試行を 4 試行，行った。音声の知覚や印象を尋ねるセクションでは，質問項目への回答を，PC のテンキーで行うことを求めた。話者の年齢推定を尋ねるセクションでは，質問項目への回答を，質問紙に記入することを求めた。音声の知覚や印象を尋ねるセクションにおける質問項目の呈示順序は，3 通り存在した。話者の年齢推定を尋ねるセクションでは，PC 画面上に常に，今，音声がいくつ流れたのかを呈示した。

両セクション終了後，参加者に謝礼としてお菓子を配布し，内観報告を求めて，実験を終了した。実験全体の所要時間は約 30 分であった。

結　果

得点化

得点化を行うにあたり，参加者の回答を確認した。その結果，不真面目に回答したと見なせる回答は存在し

なかった。そのため，参加者 21 名全員の回答を，有効回答とした。

速さ感・高さ感・自然性を測定するために用いた「遅い-速い」「低い–高い」「不自然な-自然な」に対して，参加者に 7 件法で回答を求めていた。これらの項目においては，1 から 7 の数字をそのまま得点とし，それぞれ速さ感得点，高さ感得点，自然性得点とした。これらの得点は，理論的には 1 点から 7 点に分布し，得点が高いほど，発話速度が速く，声の高さが高く，音声が自然であると参加者が判断したことを示す。話者の年齢推定に対しては，参加者に質問紙に整数を記入することで回答を求めていた。話者の年齢推定については，質問紙に記入された整数をそのまま得点とした。悠長さを測定するために用いた「ゆっくり話す–早口な」「せっかちな–ゆったりとしている」「ぞんざいな–丁寧な」に対して，参加者に 7 件法で回答を求めていた。1 から 7 の数字をそのまま得点とし，橋本・古屋（2019）に従って悠長さ得点を算出した。悠長さ得点は，理論的には 1 点から 7 点に分布し，得点が高いほど，悠長であると参加者が判断したことを示す。それぞれの得点の平均値と標準偏差を Table 1 に示す。

Table 1
各得点の平均値と標準偏差

	速さ感		高さ感		自然性		話者の年齢推定		悠長さ	
条件	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
58.14%	6.21	0.84	4.62	0.79	3.33	1.47	22.88	5.39	2.34	0.97
60.98%	6.19	0.71	4.43	0.64	3.52	1.15	22.45	5.35	2.44	0.85
64.10%	5.79	0.87	4.21	0.66	3.81	1.20	22.24	4.47	2.86	0.99
67.57%	5.57	0.86	4.38	0.63	4.43	1.12	23.48	5.16	2.96	0.82
71.43%	5.45	0.71	4.21	0.64	4.50	1.52	22.86	4.26	3.07	0.81
75.76%	5.00	0.83	4.26	0.49	5.05	1.01	23.26	4.90	3.60	0.88
80.65%	4.81	0.89	4.17	0.48	5.62	0.93	23.31	4.38	3.99	0.90
86.21%	4.17	0.73	4.19	0.40	5.93	0.75	23.29	4.72	4.29	0.83
92.59%	4.00	0.62	4.12	0.47	5.90	0.85	23.31	3.87	4.72	0.56
100.00%	3.69	0.78	3.95	0.50	5.60	1.03	23.31	4.14	4.82	0.77
108.00%	3.55	0.99	4.17	0.48	4.95	1.07	23.36	6.04	5.07	0.59
116.00%	3.07	0.87	4.07	0.60	4.45	1.15	23.79	7.28	5.22	0.58
124.00%	2.81	0.83	4.00	0.52	4.14	1.26	24.14	8.60	5.33	0.63
132.00%	2.38	0.82	3.90	0.49	3.64	1.23	23.36	8.64	5.35	0.76
140.00%	2.24	0.62	3.86	0.53	3.31	1.29	24.50	11.47	5.48	0.64
148.00%	2.17	1.01	3.90	0.70	3.00	1.00	23.48	11.41	5.44	0.76
156.00%	1.88	0.89	3.81	0.77	2.48	1.32	22.31	10.43	5.60	0.68
164.00%	1.62	0.85	3.76	0.77	2.17	1.06	22.40	10.94	5.71	0.65
172.00%	1.43	0.59	3.76	0.58	2.05	1.17	23.55	10.14	5.83	0.63

回帰分析

持続時間要因に対して対数変換を行った後，回帰分析による 1—3 次式の推定を行った。当てはまりの良さの判断には，*AIC*（Akaike's Information Criterion）と，R（ver.3.5.1）の anova 関数を用いた。*AIC* は，数値が低い方が当てはまりが良いという判断になる。anova 関数は，1 次式と 2 次式，および，2 次式と 3 次式の間の比較を行い，高次の式の方が当てはまりが良ければ有意な差が認められ，当てはまりが良いという判断になる。なお，高次式の方が当てはまりが良くなる（データの複雑な変化に合った式になりやすい）という性質がある

ため，1 次式と 2 次式の間に有意な差が認められなければ，2 次式と 3 次式の間に有意な差が認められたとしても，1 次式に当てはまりが良いと判断する。

速さ感　1—3 次式の *AIC* はそれぞれ，1938.979，1937.029，1939.018 であった。すなわち，1—3 次式の中で最も *AIC* が小さかった式は，2 次式であった。また，anova 関数を用いた分析では，1 次式と 2 次式の間には有意な差が認められた（$p = .047$）が，2 次式と 3 次式の間には有意な差が認められなかった（$p = .917$）。したがって，2 次式が最も当てはまりが良いと判断した。R^2 は.789，調整済み R^2 は.788 であった（$F(2, 795) = 1484.000$, $p < .001$）。速さ感の散布図および回帰式を Figure 1 に示す。

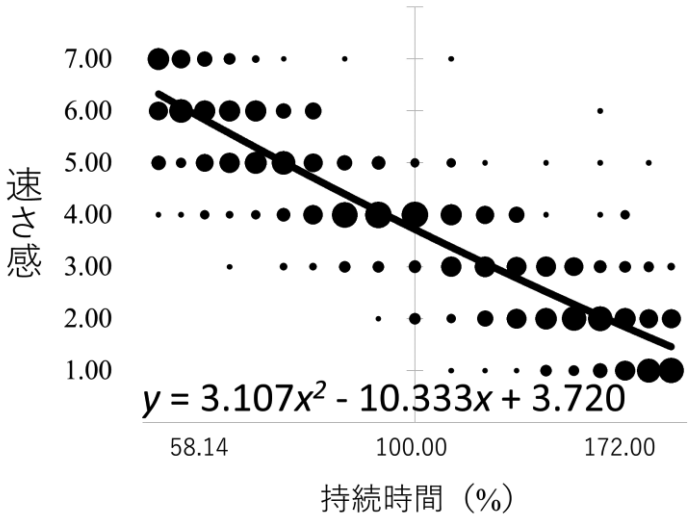

Figure 1.　速さ感の散布図および回帰式。ドットはデータ数を示す。ドットが大きいほどデータ数が多い。

高さ感　1—3 次式の *AIC* はそれぞれ，2249.076，2250.910，2251.819 であった。すなわち，1—3 次式の中で最も *AIC* が小さかった式は，1 次式であった。また，anova 関数を用いた分析では，1 次式と 2 次式，および，2 次式と 3 次式の間には，有意な差は認められなかった（$p = .685$; $p = .298$）。したがって，1 次式が最も当てはまりが良いと判断した。R^2 は.046，調整済み R^2 は.046 であった（$F(1, 796) = 38.140$, $p < .001$）。高さ感の散布図および回帰式を Figure 2 に示す。

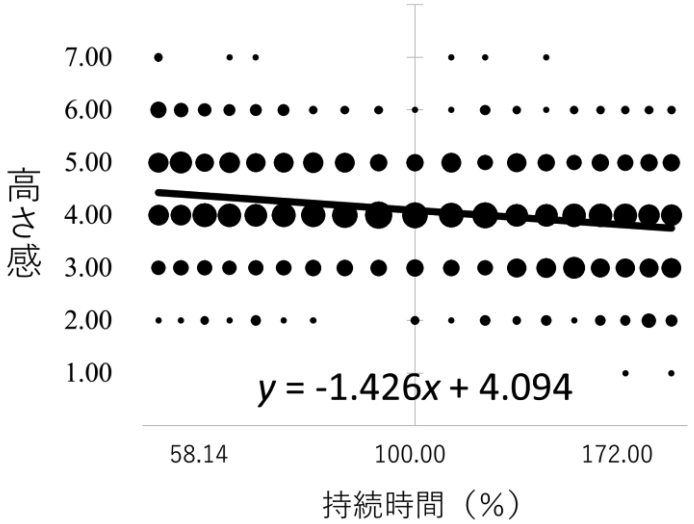

Figure 2.　高さ感の散布図および回帰式。ドットはデータ数を示す。ドットが大きいほどデータ数が多い。

自然性　1—3 次式の *AIC* はそれぞれ，3164.283，2868.472，2852.645 であった。すなわち，1—3 次式の中で最も *AIC* が小さかった式は，3 次式であった。また，anova 関数を用いた分析では，1 次式と 2 次式，および，2 次式と 3 次式の間に有意な差が認められた（$p < .001$; $p < .001$）。したがって，3 次式が最も当てはまりが良いと判断した。R^2 は.396，調整済み R^2 は.394 であった（$F(3, 794) = 173.400$, $p < .001$）。自然性の散布図および回帰式を Figure 3 に示す。

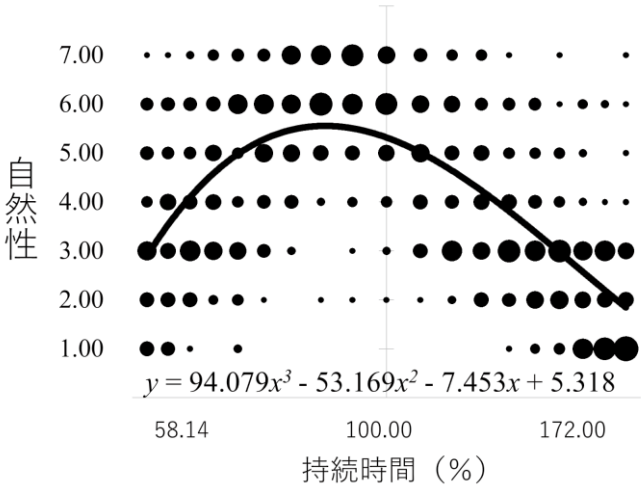

Figure 3.　自然性の散布図および回帰式。ドットはデータ数を示す。ドットが大きいほどデータ数が多い。

話者の年齢推定　1—3 次式の *AIC* はそれぞれ，5458.459，5459.314，5461.131 であった。すなわち，1—3 次式の中で最も *AIC* が小さかった式は，3 次式であった。また，anova 関数を用いた分析では，1 次式と 2 次式，および，2 次式と 3 次式の間には，有意な差は認められなかった（$p = .286$; $p = .670$）。加えて，1 次式自体が有意ではなかった（$F(1, 796) = 0.486$, $p = .486$）。したがって，話者の年齢推定については，当てはまりの良い回帰式はないと判断した。

悠長さ　1—3 次式の *AIC* はそれぞれ，1938.647，1843.332，1845.326 であった。すなわち，1—3 次式の中で最も *AIC* が小さかった式は，2 次式であった。また，anova 関数を用いた分析では，1 次式と 2 次式の間には有意な差が認められたが，2 次式と 3 次式の間には有意な差は認められなかった（$p < .001$; $p = .936$）。したがって，2 次式が最も当てはまりが良いと判断した。R^2 は.697，調整済み R^2 は.696 であった（$F(2, 795) = 912.600$, $p < .001$）。悠長さの散布図および回帰式を Figure 4 に示す。

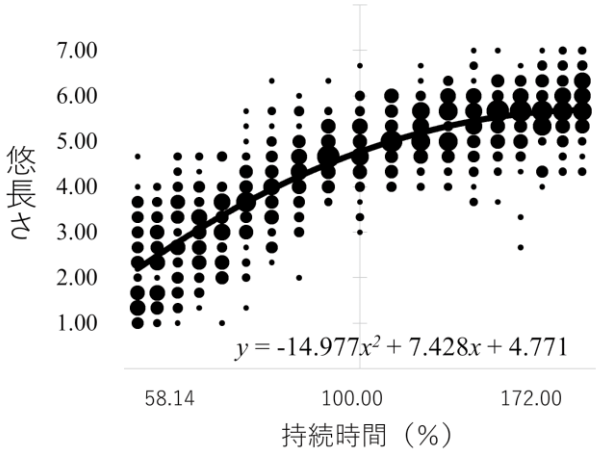

Figure 4.　悠長さの散布図および回帰式。ドットはデータ数を示す。ドットが大きいほどデータ数が多い。

t 検定

速さ感・高さ感・悠長さに対する *t* 検定を行った。まず，自然性の 3 次式（$y = 94.079x^3 - 53.169x^2 - 7.453x + 5.318$）を用いて，自然性得点の中央値である 4.00 点になる持続時間を算出した。その結果，2.13 s（本研究の条件で言えば 63.51%条件）と 4.32 s（本研究の条件で言えば 128.45%条件）となった。続いて，2.13 s よりも持続時間が長くて 2.13 s に最も持続時間が近い 64.10%条件（2.15 s）と，4.32 s よりも持続時間が短くて 4.32 s に最も持続時間が近い 124.00%条件（4.17 s）の *t* 検定を，速さ感・高さ感・悠長さに対して行った。なお，64.10%条件と 124.00%条件の得点は，速さ感では 5.79 と 2.81，高さ感では 4.21 と 4.00，悠長さでは 2.86 と 5.33 であった。*t* 検定の結果，速さ感には有意な差が認められた（$t(41) = 15.321, p < .001$）。すなわち，62.10%条件の方が 124.00%条件よりも速さ感の得点が高かった。高さ感には有意な差が認められなかった（$t(41) = 1.502, p = .141$）。悠長さにも有意な差が認められた（$t(41) = -11.870, p < .001$）。すなわち，124.00%条件の方が 62.10%条件よりも悠長さの得点が高かった。

考　察

本研究の目的は，発話速度に変化があったと感じながらも不自然ではない持続時間の範囲を知り，その持続時間の範囲の中でパラ言語知覚と推測の議論を行うことであった。発話速度に変化があったと感じながらも不自然ではない持続時間の範囲を知るために，速さ感・自然性を測定した。パラ言語知覚の議論のために，高さ感を測定した。推測の議論のために，話者の年齢推定・悠長さを測定した。

当てはまりの良い回帰式

回帰分析の結果，速さ感と悠長さが 2 次式に，高さ感が 1 次式に最も当てはまりが良かった。自然性は 3 次式に最も当てはまりが良かった。話者の年齢推定は当てはまりの良い回帰式を得ることができなかった。

速さ感で最も当てはまりの良かった 2 次式を基に，本研究の実験音声について述べる。速さ感の 2 次式（$y = 3.107x^2 - 10.333x + 3.720$）を用いて，本研究における速さ感の得点の中央値である，4.00 点となる持続時間を算出すると，3.16 s になった（本研究の条件で言えば 94.00%条件）。実験の結果として得られた得点も，92.59%条件が 4.00 点となった。したがって，本研究の本試行で用いた実験音声は，遅いと知覚される音声の方がやや多かった。

遅いと知覚される音声が多かったことによる他の測定項目への影響を，高さ感・悠長さ・自然性のそれぞれで最も当てはまりの良かった次数の式を基にして述べる。高さ感の 1 次式（$y = -1.426x + 4.094$）を用いて，4.00 点となる持続時間を算出すると，3.91 s（本研究の条件で言えば 116.39%条件）となる。悠長さの 2 次式（$y = -14.977x^2 + 7.428x + 4.771$）を用いて，同様に 4.00 点となる持続時間を算出すると，2.74 s（本研究の条件で言えば 81.63%条件）となる。自然性の 3 次式（$y = 94.079x^3 - 53.169x^2 - 7.453x + 5.318$）を用いて，最も自然性得点が高くなる持続時間を算出すると，2.92 s（本研究の条件で言えば 87.02%条件）となる。これらの計算結果から，発話速度が遅いと知覚される音声が多かったことにより，高さ感と悠長さの得点が中央値の 4.00 となる条件や，自然性の得点が最も高くなる条件も，偏っていたと述べることができる。ただし，本研究の実験音声には偏りが生じていたものの，高さ感・自然性・悠長さの全てで，発話速度が遅いと知覚される音声が多い場合に想定される偏りが見られた。したがって，本研究で得られた得点および回帰式には一定の信頼があるだろう。なお，自然性では，持続時間の偏りが生じたために持続時間の長い条件で，知覚における床効果が起こり，3 次式に当てはまりが最も良いと判断された可能性がある。

話者の年齢推定は，当てはまりの良い回帰式を得ることができなかった。すなわち，持続時間による話者の年齢推定への評価は小さいと言えるだろう。一方で，森・前川・粕谷（2014）では，異なる男性話者の分析を行い，ポーズ数・話速（モーラ/s）・平均ポーズ長（ms）といった，発話の時間に関する指標が話者の年齢推定に影響すると結論付けている。したがって，発話の時間に関する指標の中で，持続時間のみの変化による話者の年齢推定への影響は小さく，ポーズ数などの発話の時間に関する指標との複合的な違いによって，話者の年

齢推定の変化が現れる可能性があると考えられるだろう。また，人は年齢に伴い音声器官が変化し，その結果，フォルマント周波数が変化することが明らかになっている（e.g. Peterson & Barney, 1952）。話者の年齢推定を行う際には，人という生物として自然に起こる特徴の変化が，重要な手掛かりとなり得るだろう。すなわち，話者の年齢推定には，発話の時間に関する指標よりもフォルマント周波数の方が重要視されている可能性がある。加えて，平均基本周波数で同様の検討を行った橋本（2020）においては，回帰式を得ることができた。したがって，話者の年齢推定については，発話の時間に関する指標だけでなく，フォルマント周波数や平均基本周波数といった要素との関連にも注目しながら，今後さらに検討を行う必要があるだろう。

ここで，最も当てはまりの良い回帰式について，先行研究と比較した考察をする。本研究では，速さ感は 2 次式に，自然性は 3 次式に最も当てはまりが良かった。発話している時間のみに注目した場合ではあるが，内田（2005）では，速さ感も自然性も 2 次式に最も当てはまりが良かった。遅いと知覚される音声が多かったため積極的に議論することはできないが，速さ感は内田（2005）を追従し，自然性は内田（2005）とは異なる結果となった。このように，回帰式について，先行研究と比較した考察を行えるが，当てはまりの良い回帰式は，切り取る持続時間によって異なるだろう。例えば，速さ感については，本研究も内田（2005）も 2 次式に最も当てはまりが良かった。しかし，もっと持続時間の短い，あるいは，長い音声を実験に用いれば，本研究の自然性の結果と同じように，どこかで知覚における床効果や天井効果が現れ，最も当てはまりの良い回帰式の次数も変化するだろう。特に本研究では，発話速度に変化があったと感じながらも不自然ではない持続時間の範囲を知るという目的のために，全てが自然でないことも想定されていた。最も当てはまりの良い次数については，今後も慎重に検討する必要があるだろう。

発話速度に変化があったと感じながらも不自然ではない持続時間の範囲とその範囲内で変化する評価

まず，発話速度の変化があったと感じながらも，音声の自然性が保たれる持続時間の範囲について述べる。自然性の 3 次式（$y = 94.079x^3 - 53.169x^2 - 7.453x + 5.318$）を用いて，自然性得点の中央値である 4.00 点になる持続時間を算出した。その結果，4.00 点になる持続時間は，2.13 s（本研究の条件で言えば 63.51%条件）と 4.32 s（本研究の条件で言えば 128.45%条件）であった。モーラ/ s で表すと，10.33 モーラ/ s から 5.09 モーラ/ s となる。次に，速さ感について，2.13 s よりも持続時間が長くて 2.13 s に最も持続時間が近い 64.10%条件（2.15 s）と，4.32 s よりも持続時間が短くて 4.32 s に最も持続時間が近い 124.00%条件（4.17 s）の *t* 検定を行った。なお，64.10%条件と 124.00%条件の速さ感の得点は，5.79 と 2.81 であった。その結果，速さ感に有意な差が認められた。速さ感について有意な差が認められたことから，64.10%条件と 124.00%条件は，発話速度に変化があったと感じながらも不自然ではない持続時間であると言える。モーラ/ s で表すと，10.23 モーラ/ s から 5.28 モーラ/ s となる。

次に，発話速度に変化があったと感じながらも不自然ではない持続時間の範囲の中における，高さ感と悠長さの考察を行う。高さ感と悠長さについて，速さ感と同様に，64.10%条件と 124.00%条件の *t* 検定を行った。なお，64.10%条件と 124.00%条件の得点は，高さ感では 4.21 と 4.00，悠長さでは 2.86 と 5.33 であった。その結果，悠長さは有意な差が認められたが，高さ感には有意な差が認められなかった。したがって，有意な差が認められた悠長さは，自然な範囲内の持続時間で変化することが示された。高さ感については，内田（2020）によって持続時間で変化し得る知覚であることが示されていた。しかし，本研究の *t* 検定では，有意な差が認められなかった。また，本研究の高さ感の 1 次式の調整済み R^2 は.046 と，小さい値であった。したがって，持続時間の変化による高さ感の変化はないとは言えないが，その影響は小さく，通常はほぼ感じ取ることができないと言えるだろう。

総括

本研究では，発話速度に変化があったと感じながらも不自然ではない持続時間の範囲を示すと共に，その範囲の中で高さ感の知覚や悠長さの推測についての検討を行った。ただし，本研究にも限界点がある。本研究では発話全体の持続時間を一律に変化させたため，母音をはじめとした一部の音素が他の音素に比べてより大き

く伸縮する，自然音声の変化に対する聞こえ方については，不明である。持続時間を 100%条件から対数倍になるように時間伸縮率を変化させていることにも注意が必要である。また，自然性の回答の際，人の音声としての自然性を指しているのか，合成音声としての自然性を指しているのか，参加者によって解釈が異なっていた可能性がある。今後は，音声の再合成方法，設問の設置方法に，より留意した検討が必要になるだろう。

本研究と橋本（2020）によって，持続時間と平均基本周波数の自然性を保ちつつ，発話速度や声の高さが変化したと知覚できる合成音声作成の指標の一つが示された。この指標を基にして，本研究で扱った悠長さ以外についての再検討も可能だろう。また，橋本・古屋（2019）のように性格特性を測定し，持続時間と平均基本周波数の自然性を保ちつつ，変化したと知覚できる音声の中での特性推論の過程も明らかにできるだろう。加えて，パラ言語には声の大きさや抑揚など，自然かつ変化の知覚の起こる音声の合成範囲の検討がなされていないパラ言語がある。今後，更に多くのパラ言語に対して，同様の検討を続けることも必要になるだろう。

引用文献

Allport, G., W., & Cantril, H. (1934). Judging personality from voice. *Journal of Social Psychology*, *5*, 37-55.

Apple, W., Streeter, L. A., & Krauss, R. M. (1979). Effects of pitch and speech rate on personal attributions. *Journal of personality and social psychology*, *37*, 715-727.

Asch, S. E. (1946). Forming impressions of personality. *The journal of abnormal and social psychology*, *41*, 258-290.

Boersma, P., & Weenink, D. (2015). *Praat: doing phonetics by computer.* Retrieved from http://www.fon.hum.uva.nl/praat/ (August 1, 2015.)

大坊 郁夫（1998）．しぐさのコミュニケーション――人は親しみをどう伝えあうか――　サイエンス社

藤原 武弘（1986）．態度変容と印象形成に及ぼすスピーチ速度とハンドジェスチャーの効果　心理学研究，*57*，200-206.

橋本 和奈実（2020）．話者の平均基本周波数は聴者を揺さぶる――知覚・年齢推定・気分に与える影響――　法政大学大学院紀要，*84*，17-27.

橋本 和奈実・古屋 健（2019）．発話速度と声の高さが特性推論に及ぼす影響――二段階推論仮説に基づいて――　応用心理学研究，*45*，15-25.

笠原 洋子・厳島 行雄（2007）．声の認識――声の様態が年齢推定に与える影響――　法と心理，*6*，71-84.

丸島 歩（2008）．発話速度の認知に関する一考察――基本周波数の変動との関連性に注目して――　言語学論叢　オンライン版，*1*(27)，70-85.

森 大毅・前川 喜久雄・粕谷 英樹（2014）．音声は何を伝えているか――感情・パラ言語情報・個人性の音声科学――　コロナ社

Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of the vowels. *The journal of the acoustical society of America*, *24*, 175-184.

Scherer, K. R., Banse, R., Wallbott, H. G., & Goldbeck, T. (1991). Vocal cues in emotion encoding and decoding. *Motivation and emotion*, *15*, 123-148.

内田 照久（2000）．音声の発話速度の制御がピッチ感及び話者の性格印象に与える影響　日本音響学会誌，*56*，396-405.

内田 照久（2002）．音声の発話速度が話者の性格印象に与える影響　心理学研究，*73*，131-139.

内田 照久（2005）．音声中の抑揚の大きさと変化パターンが話者の性格印象に与える影響　心理学研究，*76*，382-390.