社団法人　電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
TECHNICAL REPORT OF IEICE.

基本周波数のシフトが個人性知覚に及ぼす影響

北村　達也†　　川元　広樹†

† 甲南大学知能情報学部　〒 658–8501 兵庫県神戸市東灘区岡本 8–9–1
E-mail: †t-kitamu@konan-u.ac.jp

あらまし　男性話者 5 名の文音声を対象にして基本周波数のシフトが個人性知覚に与える影響を調査した．オリジナルの基本周波数に対して，−6 半音から+6 半音まで 2 半音ずつシフトさせた分析合成音を刺激音とした．その刺激音をオリジナルの基本周波数をもつ刺激音と対で実験協力者に提示し，2 つの刺激音の話者が同一か否かを回答させた．その結果，オリジナルの基本周波数から ±2 半音のシフトは話者識別への影響が小さいことが明らかとなった．この知覚特性は，声の高さの判定に関する知覚特性とは異なるものであった．

キーワード　個人性，基本周波数，話者識別

Effects of Shift of Pitch Frequency on Perception of Speaker Individualities

Tatsuya KITAMURA† and Hiroki KAWAMOTO†

† Faculty of Intelligence and Informatics, Konan University　Okamoto 8–9–9, Higashinada, Kobe, Hyogo, 658–8501 Japan
E-mail: †t-kitamu@konan-u.ac.jp

Abstract　This study investigated effects of the shift of the pitch frequency of sentence speech uttered by five male speakers for perceptual speaker identification. Stimuli used in experiments were re-synthesized speech, of which the pitch frequency was shifted from −6 to +6 semitones in increments of 2 semitones. In the experiments, participants were asked to judge whether the speakers of the stimuli were the same or not. The results showed that the 2-semitone-shift of the pitch frequency does not affect speaker identification. This perception characteristic is difficult from that for the pitch height of the sentence speech.

Key words　Speaker individualities, Pitch frequency, Speaker identification

1. はじめに

音声中の特徴量は発話のたびに変動するにもかかわらず，その個人性は安定して知覚される．また，よく似た声をもつ 2 名の音声であっても特徴量が全く同じわけではなく [1]，目標話者とそっくりに聞こえる物真似音声でも音響的特徴量には差異がある [2]．従って，音声の個人性に関する脳内表現においては，確率的な変動や若干の差異を許容する形で音響的特徴量と話者とが対応付けられているはずである．

この許容範囲を把握することができれば，個人性知覚メカニズムの解明に寄与することができる．そればかりでなく，知覚上の話者間距離を定義できるため話者認識技術の進歩にも貢献すると考えられる．そこで，本研究では手始めに文音声の基本周波数の周波数方向のシフトを対象にして，個人性知覚上の許容範囲を調査する．

基本周波数の個人性知覚への寄与については以前から研究されてきた．伊藤と斉藤 [3] は文音声を対象にして各種音響的特徴量と個人性知覚との関係を調査した．その結果，スペクトル包絡特性が保持されている条件では基本周波数とテンポは話者の識別にあまり影響を与えないが，スペクトル包絡特性が保持されていない条件では基本周波数とテンポの寄与が大きくなると報告した．本研究は，スペクトル包絡特性が保持されている条件下で，どの程度の平均基本周波数の変化が許容されるのかを明らかにする．

また，Akagi & Ienaga [4] は基本周波数の動的成分と静的成分の個人性知覚への寄与について調べた．基本周波数のいわゆる藤崎モデルを用いて静的成分と動的成分を独立に制御し，個人性知覚には後者の寄与が大きいことを示した．本研究では動的成分は固定した上で静的成分 (平均基本周波数) の変化の影響を調査する．

橋本ら [5] は文音声を対象にしてスペクトル，基本周波数，音素継続時間と個人性知覚との関係を報告している．そして，スペクトルと基本周波数は個人性知覚に寄与し，その寄与度は音響的特徴量の差異に依存することを示した．

これに関連する研究として Izumida & Kitamura [1] がある．彼らは文音声を対象にして話者間の音素継続時間長と基本周波数の置換が話者識別に与える影響を調査した．その結果，2 話者の識別に寄与する音響的特徴量は，これらの話者の音声における音響的特徴量の差異に依存することを示した．すなわち，多くの場合，音素継続時間長と基本周波数は個人性知覚に寄与せず，スペクトル包絡における差異が重要となるが，スペクトル包絡における個人性情報が類似している 2 話者においては，音素継続時間長と基本周波数における差異が重要となる．

ここで問題となるのは，2 話者の音声の音響特徴量における差異が知覚上いかに計算されているのかということである．出水田と赤木 [6] は，基本周波数の変化の傾きとスペクトル傾斜の変動幅と個人性知覚の関係について興味深い結果を報告している．彼らは，これらの音響的特徴量を制御することによって，「はきはき」の印象を制御できることを示した．そして，「はきはき」の程度を変化させると，ある程度までは個人性に変化がないが，ある点を超えると急に大きな変化が生じることを示した．これは，個人内で生じうる音響的特徴量のある程度の変動に対しては感度が低くなっており，それ以上の差異がある場合には異なる話者と識別するという，いわばカテゴリカルな知覚メカニズムが個人性知覚にも存在することを示唆している．

そこで，本研究では制御の容易な基本周波数のシフトを対象にして，個人性知覚に同様の特性が見られるか否かを調査する．実験においては，成人男性 5 名の文音声を対象にして，−6 半音から+6 半音まで基本周波数を対数軸上でシフトさせた刺激音を用いて，個人性知覚特性を調べる．さらに，この知覚特性と声の高さの知覚特性の比較を行う．

2. 実　験　1

実験 1 では文音声の基本周波数のシフトが個人性知覚に及ぼす影響を調査するための聴取実験を行った．

2.1 実 験 条 件

2.1.1 刺　激　音

原音声は，ATR 音声データベースセット C に含まれる関東出身男性話者 5 名 (M318，M509，M601，M603，M710) の文音声である．これらの話者の年齢は 21 歳から 37 歳で，本研究の実験協力者にとって未知話者である．刺激音としては，(a)「あの坂を上れば海が見える」，(b)「飛ぶ自由を得ることは人類の夢だった」の 2 文を用いた．なお，この音声データベースは高度言語情報融合フォーラム (ALAGIN) によって公開されている．

上記の原音声の基本周波数全体を対数軸上でシフトさせたものを刺激音とした．刺激音の作成には STRAIGHT [7] を用いた．文全体の基本周波数を十二平均律に基づき −6 半音から+6 半音まで 2 半音ずつシフトさせ，基本周波数が変化していないものも含めて 7 種類の刺激音を作成した．一例として，オリジナルの周波数を 120 Hz としたときに −6 半音から+6 半音まで 2 半音ずつシフトさせた周波数を表 1 に示す．

表 1　オリジナルの周波数を 120 Hz としたときの −6 半音から+6 半音までの周波数

shift [semitone]	frequency [Hz]
+6	169.7
+4	151.2
+2	134.7
0	120.0
−2	106.9
−4	95.2
−6	84.9

刺激音の標本化周波数は 20 kHz，量子化ビット数は 16 bit である．また，刺激音の最大振幅を話者間で正規化した．

2.1.2 実験協力者

21 歳から 23 歳の聴覚に異常のない男性 16 名，女性 4 名の計 20 名が参加した．

2.1.3 実 験 方 法

実験では，原音声の基本周波数をもつ刺激音と上記の 7 種の刺激音のうちの 1 つから成る刺激対を実験協力者に提示した．2 つの刺激音の話者は同一である．順序効果を排除するために順序を入れ替えた刺激音対も提示し，1 つの刺激音対は 2 度提示した．従って，試行数は 1 文につき (7 種の刺激音)×(2 順序)×(2 回)×(5 話者)=140 である．

実験は防音室にて Praat を用いて実施した [8] [9]．実験協力者は，刺激対の話者が同一話者か異なる話者かを PC の画面上に表示されたボタンをクリックすることにより回答した．刺激音は 1 度だけ聞き直すことを許した．実験は刺激音の話者ごとに 2 セットに分けて実施し，間に数分間の休憩を入れた．

刺激音は，PC から出力された音声をヘッドフォンアンプ (Fostex HP-A3) にて D/A 変換し，密閉型ヘッドフォン (Sennheiser HDA200) にて提示した．実験協力者は各自の聴きやすいレベルで聴取した．

2.2 結　　果

文 (a)，文 (b) に関して同一話者と回答された割合をそれぞれ図 1，図 2 に示す．これらの結果は実験協力者間で平均した値である．エラーバーは標準偏差を表す．

基本周波数のシフトにより結果に差異があるか否かを評価するため有意水準 5 %の分散分析を行った．その結果，文 (a) で $F(6, 133) = 68.38$，文 (b) で $F(6, 133) = 52.38$ となり，いずれの文についても有意差があることがわかった．さらに，Tukey の HSD 検定により多重比較を行ったところ，文 (a)，文 (b) ともすべての群の間に有意差があった．

図 1，図 2 から，基本周波数の ±2 半音のシフトは個人性知覚への影響が小さく，それ以上のシフト量は個人性知覚への影響が大きいことがわかる．基本周波数を ±2 半音シフトさせても同一話者と回答された割合は 90 %以上を保っている．それに対して，±4 半音のシフトにより同一話者と回答された割合はほぼチャンスレベルとなり，±6 半音のシフトにより 30 %程

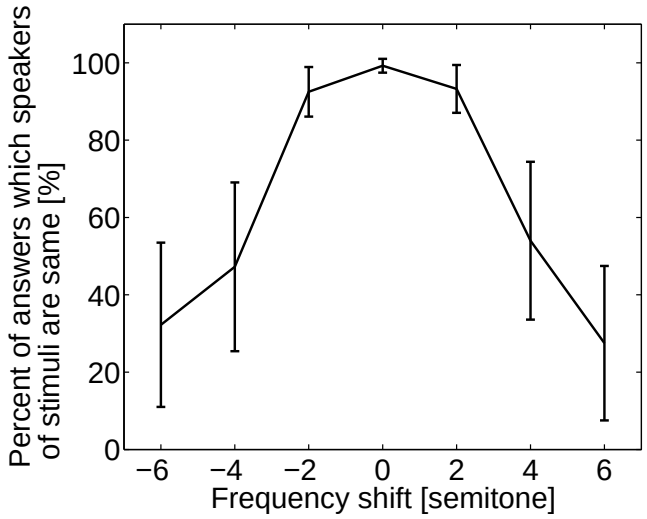

図 1　オリジナルの基本周波数をもつ文 (a) の刺激音 (Frequency shift=0 semitone) との比較において同一話者と回答された割合 (%)

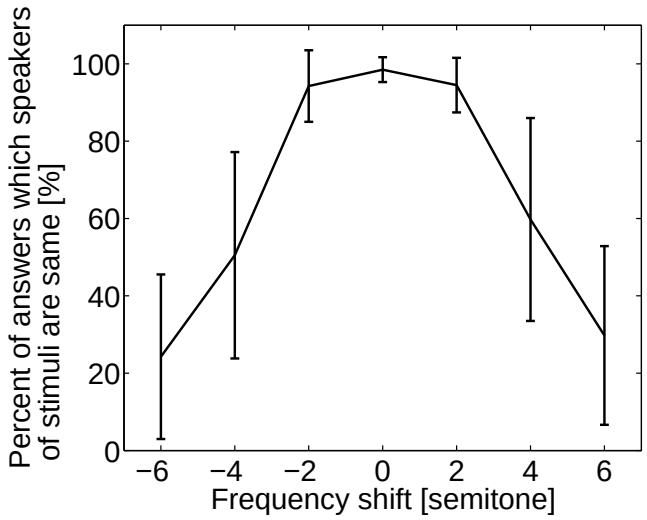

図 2　オリジナルの基本周波数をもつ文 (b) の刺激音 (Frequency shift=0 semitone) との比較において同一話者と回答された割合 (%)

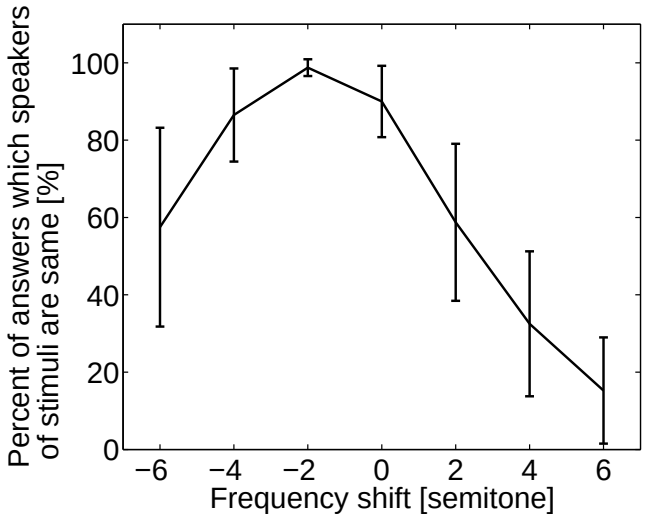

図 3　2 半音下降させた基本周波数をもつ文 (a) の刺激音 (Frequency shift=−2 semitone) との比較において同一話者と回答された割合 (%)

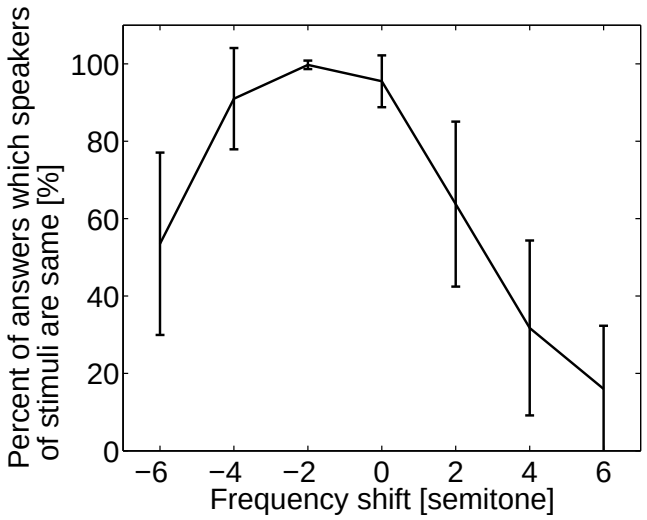

図 4　2 半音下降させた基本周波数をもつ文 (b) の刺激音 (Frequency shift=−2 semitone) との比較において同一話者と回答された割合 (%)

度まで減少する．

また，基本周波数の上昇と下降による影響はほぼ等しく，シフトによる影響は変化方向に対して対称性がある．

3. 実　験　2

実験 1 により，基本周波数のシフトが ±2 半音以内であれば個人性知覚への影響が小さいことが明らかとなった．そこで，±2 半音という範囲が成人男性の平均的な基本周波数に対して相対的に定まっているものなのか，もしくは比較対象の基本周波数に対して相対的に定まるものなのかを調査する．このために，基準となる音声の基本周波数を 2 半音下げて実験 1 と同様の聴取実験を行った．

もし前者が正しければ，基準となる音声の基本周波数にかかわらず実験 1 と同様の結果が得られるはずである．一方，後者が正しければ，−2 半音から ±2 半音以内は基本周波数シフトの影響が小さく，それ以上では大きくなるはずである．

3.1 実 験 条 件

3.1.1 刺　激　音

実験 1 と同じ刺激音を用いた．

3.1.2 実験協力者

19 歳から 22 歳の聴覚に異常のない男性 17 名，女性 3 名の計 20 名が参加した．これらの実験協力者は実験 1 と異なる．

3.1.3 実 験 方 法

実験では，原音声から基本周波数を 2 半音下げた刺激音と 7 種の刺激音のうちの 1 つから成る刺激対を実験協力者に提示した．その他の条件は実験 1 と同じである．

3.2 結　　果

文 (a)，文 (b) に関して声の高さが同一と回答された割合を

それぞれ図 3，図 4 に示す．これらの結果は実験協力者間で平均した値である．エラーバーは標準偏差を表す．

基本周波数のシフトにより結果に差異があるか否かを評価するため有意水準 5 %の分散分析を行った．その結果，文 (a) で $F(6,133) = 69.77$，文 (b) で $F(6,133) = 71.35$ となり，いずれの文についても有意差があることがわかった．さらに，Tukey の HSD 検定により多重比較を行ったところ，文 (a)，文 (b) ともすべての群の間に有意差があった．

図 3，図 4 から，実験 1 と同様に，基本周波数の ±2 半音のシフトは個人性知覚への影響が小さく，それ以上のシフト量は個人性知覚への影響が大きいことがわかる．さらに，実験 1 と同様に基本周波数の上昇と下降による影響はほぼ等しく，シフトによる影響は変化方向に対して対称性がある．従って，±2 半音という範囲は，比較対象の基本周波数に対して相対的に定まるものであることが明らかとなった．

4. 実　験　3

実験 1 および 2 において基本周波数の ±2 半音シフトは個人性知覚への影響が小さいという結果が得られた．この結果は実験協力者が基本周波数の ±2 半音シフトによるピッチの変化を検出できなかったことによる可能性がある．そこで，このことを確認するための聴取実験を行った．

4.1 実 験 条 件

4.1.1 刺　激　音

実験 1 と同じ刺激音を用いた．

4.1.2 実験協力者

20 歳から 22 歳の聴覚に異常のない男性 9 名，女性 3 名の計 12 名が参加した．一部の実験協力者は実験 1 および実験 2 と重複する．

4.1.3 実 験 方 法

刺激対，試行数とも実験 1 と同じである．実験協力者は，刺激対の声の高さが同一か否かを PC の画面上に表示されたボタンをクリックすることにより回答した．

4.2 結　　果

文 (a)，文 (b) に関して声の高さが同一と回答された割合をそれぞれ図 5，図 6 に示す．これらの結果は実験協力者間で平均した値である．エラーバーは標準偏差を表す．

基本周波数のシフトにより結果に差異があるか否かを評価するため有意水準 5 %の分散分析を行った．その結果，文 (a) で $F(6,133) = 68.38$，文 (b) で $F(6,133) = 52.38$ となり，いずれの文についても有意差があることがわかった．さらに，Tukey の HSD 検定により多重比較を行ったところ，文 (a)，文 (b) ともすべての群の間に有意差があった．

声の高さが同一と回答された割合の平均値は，± 4 半音以内の基本周波数のシフト量に対してほぼ線形に変化している．基本周波数のシフトがない条件では約 100 %であるのに対し，±2 半音シフトさせた条件で約 60 から 70 %に下降する．この結果から，全ての実験協力者が基本周波数の ±2 半音シフトによるピッチの違いを「検出できた」とまでは言えない．しかし，実験 1 で示された個人性知覚特性とは明らかに異なる特性を示している．

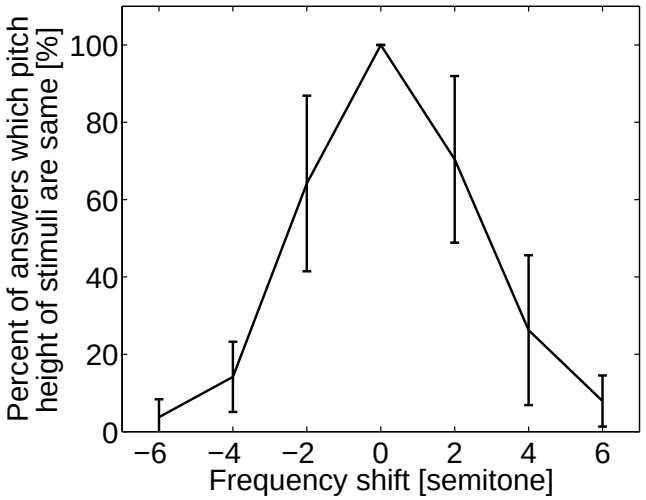

図 5　オリジナルの基本周波数をもつ文 (a) の刺激音 (Frequency shift=0 semitone) との比較において声の高さが同一と回答された割合 (%)

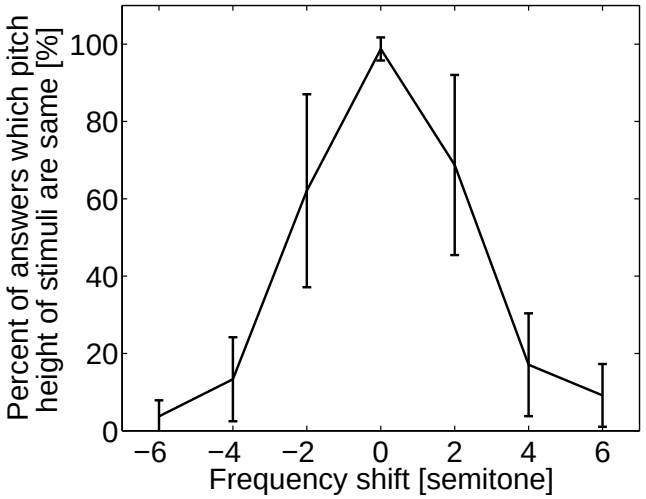

図 6　オリジナルの基本周波数をもつ文 (b) の刺激音 (Frequency shift=0 semitone) との比較において声の高さが同一と回答された割合 (%)

5. 考　　察

実験 1 によって，本研究の条件下では，文音声の基本周波数の ±2 半音のシフトは個人性知覚にほとんど影響しなかった．一方，±4 半音以上のシフトは個人性知覚への影響が大きいことが示された．以上の結果は，個人性知覚において，一定範囲内の平均基本周波数の変動は許容され，それ以上の変化は話者の特徴として認識されることを意味している．つまり，個人性知覚においても，ある種カテゴリカルな知覚特性が存在することを示唆する結果といえる．また，実験 2 によって，±2 半音という範囲は，成人男性の平均的な基本周波数に対して相対的に定まっているものではなく，比較対象の音声の基本周波数に対して相対的に定まっていることが明らかとなった．

基準となる基本周波数を 120 Hz とすると，±2 半音の変化は

106.9 Hz から 134.7 Hz までの変化に対応し，±4 半音は 95.2 Hz から 151.2 Hz までの変化に対応する (表 1)．橋本 [10] は，日本語の単語のアクセントに関する研究の中で，男性 1 名が発話した単語の母音重心点における基本周波数を計測した．その結果，母音重心点における基本周波数は非常に安定しており，標準偏差は 4.7 % 以下であることを報告した．また，本橋 [11] は『羅生門』を朗読した音声の文全体の基本周波数について調べている．彼らが対象にした 6 話者の平均基本周波数は，25 Hz から 40 Hz 程度変動した．本研究の結果から，もし平均基本周波数がこの程度安定していれば，その変動が個人性知覚に影響することはほとんどないと予想される．

一方，音声の平均基本周波数は感情や発話様式によって変化することが知られている．例えば，日本語音声を対象とした研究では，水木 [12] は無感情，驚き，喜び，怒り，恐怖，嫌悪の感情を込めて発話された「え」の音響特徴量を分析し，基本周波数は感情によって 1 オクターブ程度変化することを示している．また，宮武と匂坂 [13] は，種々の発話様式で発話された単語音声の母音中心における基本周波数を分析し，普通発話に対して −38.7 Hz から+32.7 Hz の範囲で変化すると報告した(注1)．これらの結果から，日常の音声コミュニケーションにおいては，平均基本周波数の変動が ±2 半音よりも大きくなることが頻繁に生じると考えられる．このような状況において個人性知覚が受ける影響については検討の余地がある．

また，実験 3 によって，個人性知覚に対する平均基本周波数のシフトの影響は，ピッチ知覚に対するものと異なることが明らかになった．この結果は，個人性知覚が平均基本周波数だけで決まるものではないことを示している．話者認識技術の音響的特徴量として MFCC およびその変化成分が用いられていることからも明らかな通り，スペクトル包絡には個人性情報が豊富に含まれている．個人性知覚においても，平均基本周波数よりもスペクトル包絡の寄与が大きいことが示されている [3] [5]．本研究の結果もこれに沿ったものといえる．

本研究では分析合成系を用いて平均基本周波数のみシフトさせたが，実際の発話においては基本周波数の変化に伴い声道形状も変化する [14]．基本周波数の変化に伴うスペクトル包絡の変化を考慮することによって，本研究で示した結果とは異なる結果が得られる可能性はある．

6. おわりに

本研究では，成人男性の文音声を対象にして，文全体の基本周波数のシフトが個人性知覚に及ぼす影響を調査した．その結果，±2 半音のシフトは個人性知覚にほとんど影響しないものの，±4 半音以上のシフトは個人性知覚への影響が大きいことが示された．そして，この知覚特性は，文音声のピッチに対する知覚特性とは大きく異なることが明らかになった．

本研究では，平均基本周波数のみが異なり，基本周波数の動的成分やスペクトル包絡を保持した刺激音で聴取実験を行った．今後，異なる 2 文の音声を用いて同様の聴取実験を行い，文全体の基本周波数のシフトが個人性知覚に与える影響を調査する予定である．

(注1)：意図的に声を高くもしくは低く発話した音声の基本周波数は除外してある．また，これらの値は男女各 1 名の音声から得られたものである．

謝辞　本研究の一部は平成 24 年度科研費基盤 (B)(21300071)，平成 25 年度科研費基盤 (A)(25240026) および私立大学等経常費補助金の支援を得て行われた．有益なご助言をいただきました神戸大学大学院国際文化学研究科 波多野博顕氏，グローリー (株) 出水田剛志氏に感謝いたします．

文　献

[1] T. Izumida and T. Kitamura, Study of perceptual factors for speaker identification focusing on perceptual similarity of speaker characteristics, *Acoust. Sci. & Tech.*, 32, 5, 216–219 (2011).
[2] T. Kitamura, Acoustic analysis of imitated voice produced by a professional impersonator, *Proc. Interspeech2008*, 813–816 (2008).
[3] 伊藤憲三, 斉藤収三, 音声の音響的特徴パラメータが個人性の知覚に及ぼす影響, 信学論, J65-A, 1, 101–108 (1987).
[4] M. Akagi and T. Ienaga, Speaker individuality in fundamental frequency contours and its control, *J. Acoust. Soc. Jpn.*(*E*), 18, 2, 73–80 (1997).
[5] 橋本誠, 北川敏, 樋口宜男, 音声の個人性知覚に影響を及ぼす音響的特徴の定量的分析, 音響誌, 54, 3, 169–178 (1998).
[6] 出水田剛志, 赤木正人, 音声の動的成分に着目した個人性聴取印象の検討, 音講論 (春), 423–426 (2012).
[7] H. Kawahara, I. Masuda-Katsuse and A. de Cheveigne, Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency-based F0 extraction: Possible role of a reptitive structure in sounds, *Speech Communication*, 27, 187–207 (1999).
[8] 北原真冬, 田嶋圭一, 音声分析ソフトウェア Praat を用いた聴取実験: F0 再合成による刺激作成と実験の制御, 音響誌, 67, 8, 345–350 (2011).
[9] 北原真冬, 田中邦佳, 田嶋圭一, Praat による音声加工と知覚実験の実施法, 日本音響学会第 122 回技術講習会資料 (2012).
[10] 橋本新一郎, 日本語単語アクセントの諸性質, 信学論, 56-D, 11, 654–661 (1973).
[11] 本橋幸康, ピッチの変動について：音声教材『羅生門』の分析から, 早稲田大学国語学会, 10, 75–86 (2002).
[12] 水木久美子, 感情を含む音声に関する基礎研究: 単音節の定常的解析, 人間工学, 30, 4, 225–230 (1984).
[13] 宮武正典, 匂坂芳典, 種々の発話様式に見られる韻律特徴とその制御, 信学論, J73-D-II, 12, 1929–1935 (1990).
[14] 北村達也, パーハム・モクタリ, F0 変化に伴う声道形状変化の観測, 信学技報 (EA), 104, 715, 25–28 (2005).