

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

内田照久

## 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\*

内田 照久\*<sup>1</sup>

(1999年6月25日受付)

【要旨】 話速変換音声に対する主観的なピッチ感を測定した。変換音声には平均  $F_0$  に有意差がないにも関わらず、発話速度の低下に伴って音声はより低く認知されていた。音声を 150% 伸長した場合の認知的なバイアスの推定値は、約 -14.0 mel であった。次に話速変換による話者の性格印象への影響を検討した。性格特性 5 因子モデルに基づく測定の結果、協調性は速度の低下に伴っていったん上昇、更に遅くなると低下した。外向性、開放性、誠実性では、やや速い発話にピークがあり、速度の低減と共に低下した。情緒不安定性にはあまり影響はなかった。聴取者は性格印象を特性因子ごとに多元的に評価しており、話速変換により組織的に影響を受けていた。

キーワード 話速変換, ピッチ感, 性格印象, 個人性情報, STRAIGHT 分析合成

Speech rate conversion, Pitch impression, Personality image, Individuality information, STRAIGHT synthesis by analysis

### 1. はじめに

近年、声の高さを変えることなく、また話者の個人性情報を保存したまま、音声の発話速度を変換する技術の開発が進められている<sup>1)</sup>。このような話速変換技術は、高齢者向けの放送補助機器<sup>2)</sup>や補聴器<sup>3)</sup>、更には外国語教育などへの応用が期待されている。

しかし、音声を時間軸上で操作する話速変換は、音響的な指標には現れにくい聞き手側の聴覚印象に対して、多元的な影響を及ぼしている可能性がある。放送用途や教育への応用場面で、話速を低下させて聞き手の理解度を上げようとする際に、もしそれによって話者の性格印象などが大きく変貌してしまうとしたら、それは必ずしも好ましくないことである。話速変換技術の実用化へのメドが立ちつつある現在こそ、聞き手の聴覚印象による品質評価が求められている。そして、それは感性情報処理技術の観点から、次世代の話速変換技術開発のためにも必要である。

さて、この話速変換技術は、かつてのテープ・レコーダのテープ回転速度を変化させるような方式とは大きく異なり、基本周波数 ( $F_0$ ) やスペクトル構造を保

存したまま、音声を時間軸上で伸縮することが求められている。従来の技術開発の経緯では、音響的な観点からの評価が進められており、変換音声のスペクトル歪などについては検討が行われてきた<sup>4)</sup>。

しかし、聞き手の聴覚印象の中で、話速変換音声の声の高さが本当に変化しないものとして認知されているかどうかは、聴覚心理実験を通じて確かめる必要がある。すなわち、聴取者のピッチ感 (pitch impression) による評価が求められる。そして、音響的には保存されているはずの声の高さに関して、もし聴覚印象上では変化があるように感じられるのであれば、その広義の錯覚とも考えられる認知的なバイアスについて、その錯覚量がどの程度であるかを定量的に推定しておく必要がある。

また、話速変換技術は、そのメリットを生かしながらも、話者の個人性情報は保存したまま変換することを目標としている。これまでにも、変換音声の明瞭度や聴き易さといった観点からの評価実験は行われているが、音声の中の他の情報にどのような変容が生じているかについては、必ずしも検証されていない<sup>2)</sup>。

話者の個人性情報 (individuality information) に焦点を当てて考えた場合、聞き手が音声を通じて認知する話者の性格印象 (personality image) は、言語の内容にもまして話者自身の特徴を示す大切な情報である。この話者の性格印象が、話速変換の操作の中で、どの程度保たれているのか、もしくはどのような影響を受けるのかは、放送や教育場面での応用を考え

\* Effects of the speech rate conversion on the impressions of pitch and the images of speakers' personality, by Teruhisa Uchida.

<sup>1)</sup> 大学入試センター研究開発部  
(問合せ先: 内田照久 〒153-8501 東京都目黒区駒場  
2-19-23 大学入試センター研究開発部)

の上でも、ぜひ検討しておかなければならない事項である。

ここで、話者の性格印象といった情報を扱おうとした場合、人間の性格やパーソナリティといった漠然とした対象を一定の整理された構造でとらえられなければならない、心理学的な測定に至り得ない。

近年、人間の性格の個人差をどう評価し、表現するかに関し、パーソナリティ研究の成果としてアメリカ性格心理学会で最も注目を浴びているものに、Big Fiveと呼ばれる性格特性の5因子モデルがある<sup>5)</sup>。それは、性格やパーソナリティ全般について、幅広い観点から広範な内容側面を対象としているにも関わらず、人の性格評価の次元には、各種の性格テストや評定の形式の違いを越えて、共通の安定した性格特性として5因子が見出されるとするものである。その五つの性格特性は一般に、外向性 (extraversion)、情緒不安定性 (neuroticism)、経験への開放 (openness to experience)、勤勉性 (conscientiousness)、協調性 (agreeableness) と呼ばれている。

最近、この性格特性5因子モデルに基づいた Big Five 尺度が開発された<sup>6),7)</sup>。この性格特性5因子モデルはカバーする領域が広く、更に安定した特性を示すことが知られているので、話者の性格印象の測定にあたり、このモデルに準じた尺度の利用は有効であろう。

以上をふまえて、本報告では、はじめに話速変換された音声における声の高さの印象に焦点を絞り、発話速度の操作が主観的なピッチ感に与える影響を検討する。その次に、個人性情報の保存の観点から、話者の性格印象を対象とする。そして、性格特性5因子モデルを枠組みにした測定を行い、話速変換が話者の性格印象に与える影響を検討する。

## 2. 音声の発話速度の制御が主観的なピッチ感に与える影響

実験1では、話速変換音声に対するピッチ感、及び発話速度感を指標として、評定尺度法 (rating scale method) で測定し、認知的なバイアスの生起の有無を検討する。次の実験2では、異なる話速変換方式によって変換された音声に対し、ピッチ感と発話速度感を測定し、変換方式の違いによって認知的バイアスに差異が生ずるかを吟味する。更に実験3では、音声の  $F_0$  を操作して、実際に声の高さの異なる変換音声を作成する。そして、声の高さの印象における認知バイアスを推定するため、ピッチ感の評定値と音声の  $F_0$  とを対応づけるデータの収集を試みる。

### 2.1 実験1：話速変換音声に対するピッチ感と発話速度感の評価

話速変換音声に対するピッチ感と発話速度感について、評定尺度法によって測定を行った。

#### 2.1.1 実験1の被験者

聴力健常な成人40名が参加した。年齢は21~39歳で、男性14名、女性26名である。全員とも日本語母語話者であり、出身は16都道府県に上る。

#### 2.1.2 実験1の音声刺激

音声刺激作成のため、音響学会の研究用連続音声データベース<sup>8)</sup>が使用された。文章素材として、トピックが「機械振興会館への道案内」で、対話番号 ETL 0008 R の模擬会話から、疑問文と肯定文の二つの短文が選ばれた。文章1(S1)として0005:「どういふうに行けばいいですか?」、また文章2(S2)として0033:「案内板が出ているから分かると思います。」が採用された。発声者に関しては、男声1(M1)として etl 0005 (愛知出身, 20歳代)、更に女声1(F1)として son 1003 (東京出身, 20歳代) が選定された。そして、この男女2名がそれぞれ二つの文章素材を発音したものを原音声データとした。なお、これらの音声データは、量子化16bit、サンプリング周波数16kHzで収録されたものである。

この原音声データを基に、話速変換技術 PICOLA<sup>4)</sup>により、音声持続時間を Fast 1 (80%), Original (100%), Slow 1 (125%), Slow 2 (156.25%) の4段階に操作し、話速を一様にコントロールした。

この PICOLA は TDHS<sup>9),10)</sup> の技術を起源に持つ。この技術は、本来、周波数軸上での情報圧縮を目的として、それを時間軸上での操作によって実現したものである。森田・板倉 (1986) は、この TDHS が音声持続時間の制御にも利用できることに着目し、その機能に特化したアルゴリズムとして PICOLA を開発し、話速変換量が少ない時の音質を向上させた<sup>4)</sup>。

なお、話速変換時の音声持続時間の伸縮範囲は、予備実験で自然な音声として許容された範囲を参考にした。また、変換音声の持続時間比は、順次1.25倍きざみとし、対数時間軸上で等間隔とした。

Table 1 に、変換後の各音声刺激の持続時間を示

Table 1 実験1の話速変換音声の持続時間

	80%	100%	125%	156.25%
男声1文章1	1.350	1.691	2.114	2.633
男声1文章2	1.852	2.315	2.889	3.610
女声1文章1	1.302	1.626	2.033	2.535
女声1文章2	1.889	2.360	2.947	3.683

音声持続時間 [s]

す。

2.1.3 実験1の手続き

実験試行では、二つの刺激を対として継時的に提示した。第1刺激を基準音声、第2刺激を比較音声とし、被験者は比較音声の声の高さが、先に提示された基準音声と比べてどのくらい高い(又は低い)と感じたかを15段階(-7...0...7)で評価した。また、話す速さについても、どのくらい速い(又は遅い)と感じたかを同様に15段階で評価した。なお、それぞれの刺激の前には、注意喚起のための短い1kHz純音のビーブ音が提示された。

実験は、主に防音室(理研音響製:聴力検査室AT-80 S)で行われた。音声刺激はパーソナル・コンピュータで再生され、被験者に快適なレベルの音量で、密閉型ヘッドホン(SONY:MDR-CD 270)から提示された。

刺激対は、同一比較対としてOriginal-Original(持続時間比1:1)、Slow 1-Slow 1(1:1)を、対照比較対としてFast 1-Slow 1(1:1.5625)、Original-Slow 2(1:1.5625)、Fast 1-Slow 2(1:1.953125)を使用した。

実験試行は、[2 同一比較対+(3 対照比較対×2 対内提示順序)]×2 話者×2 文章から成る、計32試行が実施された。この実験状況では、各刺激の出現回数は等しい。また、各試行の提示順序に関しては、ランダムな系列が5系列作成され、被験者8名ごとに割り当てられた。

また、主観的ピッチ感と発話速度感の評定時に、評価用紙上で両評定軸の正負の方向が固定してしまわないように、評定内容の正負方向への割り当てをおのおの反転させて組み合わせたものを使用した。すなわち、声が「高い」とする評価は、評価軸上では正方向、負方向のいずれにも割り当てが可能であり、また話すのが「速い」という評価も同様である。そこで、評価用紙では、二つの評定内容の正負方向をそれぞれ反転して組み合わせた4種を、先述の各ランダム系列ごとに作成して割り当てて、評定の特定方向への偏り、両軸間での不要な同期傾向を相殺した。

2.1.4 実験1の結果

心理実験の結果に先立って、各音声刺激の平均  $F_0$  を Table 2 に示す。算出にあたっては、音声刺激の各分析フレームごとに抽出された  $F_0$  を素データとし、その中で該当フレームが有声区間として判定されたものを平均していった。 $F_0$  抽出は、LPC 分析の残差信号の自己相関関数に基づく変形相関法で行い、分析フレーム長は男声:20 ms, 女声:10 ms で行った。

この変換音声の平均  $F_0$  に相違が見出されるかについて、各変換系列ごとに1元配置の分散分析を行った。その結果、すべての系列とも、変換系列内の四つの音声刺激間には有意な差は見出されなかった[男声1文章1:  $F_{(3,242)}=0.09$ , N.S., 男声1文章2:  $F_{(3,378)}=0.01$ , N.S., 女声1文章1:  $F_{(3,528)}=0.02$ , N.S., 女声1文章2:  $F_{(3,768)}=0.07$ , N.S.]。従って、 $F_0$  を指標にした範囲においては、話速変換音声には声の高さに関連した明らかな相違は見られなかった。

次に評定尺度法による心理実験での結果を示す。基準音声に対する比較音声の持続時間比ごとの、主観的ピッチ感、及び発話速度感の評定平均を Fig. 1 に示す。

更に被験者ごとの評定データを基に、回帰分析によって個人別のピッチ感と速度感の回帰係数と切片をそれぞれ算出した。そして、それらの平均から全体での回帰式、及びその信頼区間を求め、その結果も図中に

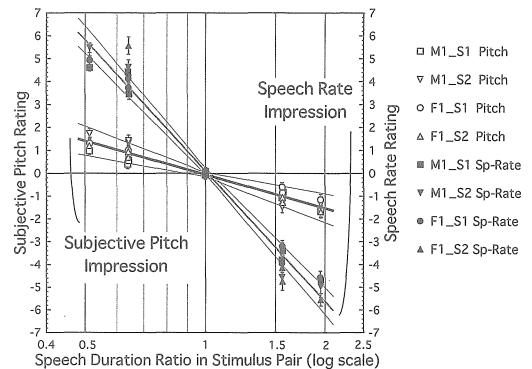


Fig. 1 話速変換音声におけるピッチ感と発話速度感基準音声に対する比較音声の時間伸縮率ごとの評定平均とS.E.、及び回帰直線と95%信頼区間。

Table 2 実験1の話速変換音声の平均  $F_0$

	80%	100%	125%	156.25%
男声1文章1	120.8 (40.0)	120.6 (38.7)	119.8 (38.1)	122.9 (39.7)
男声1文章2	159.4 (54.1)	158.8 (55.0)	158.2 (54.6)	158.7 (55.0)
女声1文章1	254.9 (64.4)	253.2 (65.5)	253.1 (64.3)	252.7 (64.6)
女声1文章2	212.6 (35.9)	211.8 (36.5)	211.2 (38.1)	211.0 (37.2)

有声区間の平均  $F_0$  [Hz] (S.D.)

示した。

図から読み取れるように、基準音声と比較音声との相対時間伸縮率が大きくなり、発話速度がゆっくりになるに従って、発話速度感の低下に留まらず、ピッチ感も同様に線形に低下していることが見出された。

従って、平均  $F_0$  においては有意差は見出されなかったものの、聴覚印象においては、話速変換による時間伸縮率と主観的なピッチ感との間に線形な関連性が観測されており、声の高さの印象においては、一定の認知的バイアスの存在が想定される。

## 2.2 実験2：異なった話速変換方式で変換した音声に対するピッチ感と発話速度感の比較

実験1では、変換音声の平均基本周波数には違いが見られないにも関わらず、話速変換による声の高さの聴覚印象への影響が観測された。これは、特定の話速変換操作によって高次のスペクトル構造などに変化が生じ、それが原因で生じた現象である可能性も考えられる。例えば、ここで用いた PICOLA では、一度圧縮した音声を伸張して再びもとの時間長にもどし、DP マッチングによって時間的な微細なずれを補正した上で LPC スペクトル歪を求めると、70%の圧縮率で1 dB 程度のスペクトル歪が観測される<sup>4)</sup>。

そこで異なった話速変換方式によって音声を処理し、それらのピッチ感と発話速度感をそれぞれ測定する。それにより、実験1で観測されたピッチ感へのバイアスが、特定の変換方式に固有の現象かどうか、また変換方式の違いによって認知的バイアスに差異が生ずるかを検証する。

### 2.2.1 実験2の被験者

実験1と同一の被験者が参加した。

### 2.2.2 実験2の音声刺激

音声データベースより、トピック「パスポート取得についての会話」、対話番号 ETL 10005 R の模擬会話から素材となる短文が選定され、文章3(S3)として、y 037:「車じゃないと、大変ですよね。」が採用された。発声者は、男声2(M2)は mit 0003 (東京出身, 30歳代), 女声2(F2)は mit 1003 (神奈川出身, 20歳代) が選ばれた。この素材を、上記の男女2名が発音したものを原音声データとした。

これらの原音声データを、PICOLA<sup>4)</sup>、NHK方式<sup>1)2)</sup>、STRAIGHT分析合成<sup>11)~13)</sup>の3種類の話速変換方式によって、実験1と同じく、それぞれ Fast 1 (80%), Original (100%), Slow 1 (125%), Slow 2 (156.25%) の4段階に変換した。

ここでNHK方式とは、音声の周期的区間では基本周期の抽出を、また非周期的区間では擬似的な基本周期の抽出を行い、後に波形周期単位での挿入削除を

Table 3 実験2の話速変換音声の持続時間

	80%	100%	125%	156.25%
PICOLA				
男声2文章3	1.395	1.751	2.183	2.733
女声2文章3	1.481	1.851	2.314	2.885
NHK方式				
男声2文章3	1.380	1.751	2.153	2.702
女声2文章3	1.461	1.851	2.278	2.852
STRAIGHT				
男声2文章3	1.401	1.751	2.189	2.736
女声2文章3	1.481	1.851	2.314	2.892
音声持続時間 [s]				

行って、時間操作を実現するものである<sup>1)2)</sup>。この技術は、放送用途での実用性を強く指向しており、DSPを用いたりリアルタイム処理を実現している。更に、より速く波形周期を抽出するため、複数窓を用いた基本周期抽出法を導入している点も特徴的である<sup>14)</sup>。更に本実験では使用しないが、映像との同期を図るため、音声の話し始めをゆっくりにし、次の息つきまでに次第に元の速さに戻すと共に、音声の“ま”の時間を短くして対応するモードも設定している<sup>1)</sup>。

なお、先述の PICOLA とこの NHK 方式は波形周期単位での操作が基本であるため、与えられた伸縮率に対して必然的に若干の時間的な誤差が生ずる。

次に STRAIGHT 分析合成は、かつての VOCODER 分析合成の持つ音声パラメータの加工の自由度を残したまま、聴覚と音声の相互作用を研究するためのツールとして新たに開発された経緯を持つ<sup>12)13)</sup>。計算量が極端に多い点が難点ではあるが、原音声に迫る自然な分析合成音を得ることができ<sup>11)</sup>。また、音声パラメータの加工が比較的容易である点が特徴的であり、教育用途への応用も試みられている<sup>15)</sup>。

この STRAIGHT 分析合成では、位相情報が意図的に破壊されるので、入力と出力の波形はかなり異なったものになる。また、スペクトルの違いも、品質から想定されるよりもはるかに大きい。しかしながら、この入力と出力の単純な違いは、聴覚的な品質の良し悪しとは必ずしも対応しないとされている<sup>11)</sup>。

Table 3 に、各変換方式で操作した音声刺激の持続時間を示す。

### 2.2.3 実験2の手続き

実験1での手続きと同様に、連続して提示される基準音声と比較音声を聞き、比較音声について、声の高さと話す速さの聴覚印象を評定した。

刺激対の組合せは、Fast 1-Slow 1 (1:1.5625)、Original-Slow 2 (1:1.5625) で、試行は{2対照比

較対×2 対内提示順序)×3 変換方式×2 話者×1 文章)から成る, 計 24 試行であった。提示方法や順序, 条件などは実験 1 と同様である。

### 2.2.4 実験 2 の結果

基準音声と比較音声の相対時間伸縮率に対する, ピッチ感と速度感の評定平均を Fig. 2 に示す。

更に, 3 種の変換方式ごとに被験者の評定データを分けた上で, 回帰分析によって個人別のピッチ感と速度感の回帰係数と切片をそれぞれ算出した。それを, あらためて変換方式ごとにまとめ直し, それぞれの平均から各変換方式ごとでの回帰式を求め, 図示した。

図より, いずれの 3 変換方式の場合でも, 実験 1 と同様, 発話速度の操作に伴うピッチ感への影響が見出された。

このピッチ感に関して, 変換方式の違いによってその影響の強さが異なるかどうかを, 被験者別の回帰分析で算出した回帰係数と切片を変数として分析した。被験者内 1 要因 (変換方式要因: 3 水準) の分散分析の結果, この 2 変数のいずれにも有意な差は見られなかった (回帰係数:  $F_{(2,78)}=2.54$ , N.S., 切片:  $F_{(2,78)}=0.64$ , N.S.)。なお, 発話速度感も同様に, 方式の違いによる有意な印象の違いは見出されなかった (回帰係数:  $F_{(2,78)}=2.56$ , N.S., 切片:  $F_{(2,78)}=0.25$ , N.S.)。

従って, 3 種の話速変換方式のいずれにおいても, ほぼ等しい水準で主観的なピッチ感へのバイアスが生じているものと考えられる。

### 2.3 実験 3: 基本周波数を変換した音声に対するピッチ感と発話速度感

実験 2 では, 変換方式の違いによるピッチ感へのバイアスに違いは見出されなかった。そこで実験 3 では, ピッチ感の評定値と音声の基本周波数 ( $F_0$ ) とを

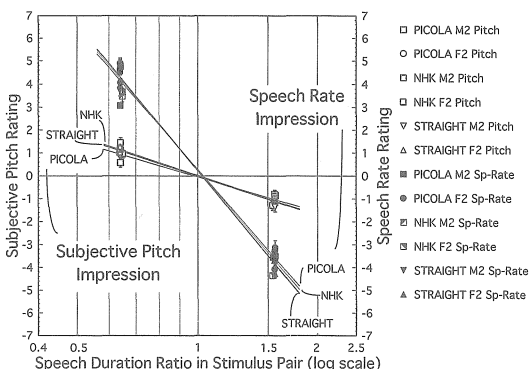


Fig. 2 3 種類の制御方式による話速変換音声に対するピッチ感と発話速度感の比較  
基準音声に対する比較音声の時間伸縮率ごとの評定平均と S.E., 及び方式別の回帰直線。

対応づけるデータの収集を試みる。STRAIGHT 分析合成を用いて音声の  $F_0$  を直接操作し, ピッチ感と発話速度感の測定を行うことにより, ピッチ感と  $F_0$  の定量的な関連づけを試みる。

#### 2.3.1 実験 3 の被験者

実験 1-2 と同一の被験者が参加した。

#### 2.3.2 実験 3 の音声刺激

素材の文章, 話者とも実験 2 と同様で, 同じ原音声データが使用された。

これらの原音声に対して, STRAIGHT 分析合成によって音声データの基本周波数を操作し, Low (90%), Original (100%), High (111.11%) の 3 段階に変換した。なお,  $F_0$  変換の範囲は, 予備実験で許容された範囲内である。また, 変換した周波数比は, 順次 1.1111 倍きざみとし, 対数周波数軸上で等間隔とした。

#### 2.3.3 実験 3 の手続き

実験 1-2 と同様の手続きで声の高さと話す速さの印象評定を行った。

刺激対は Low-Original ( $F_0$  比 1:1.1111), Original-High (1:1.1111), Low-High (1:1.2346) とし, 試行は{(3 対照比較対×2 対内提示順序)×2 話者×1 文章)から成る, 計 12 試行である。

#### 2.3.4 実験 3 の結果

心理実験結果に先立ち,  $F_0$  変換音声の分析を行った。まず, 各変換音声の平均  $F_0$  を算出した。更にこの実験は, 変換音声の音の高さと, 主観的なピッチ感といった心理的な属性との対応づけが目的であるので, 音の高さを扱う指標として心理尺度として構成された mel scale<sup>16)</sup> を用いた。

平均  $F_0$  の算出は実験 1 での方法と同様である。音声刺激の各分析フレームごとの  $F_0$  を素データとし, その中で有声フレームのデータを平均したものを平均  $F_0$  とした。更に平均ピッチの算出にあたっては, まず有声フレームのそれぞれの  $F_0$  値を, (1) 式の近似式<sup>17)</sup> を用いて mel scale 上に変換した。そのようにして得られた mel 値の平均を, その音声刺激の平均ピッチとした<sup>18)</sup>。

$$m[\text{mel}] = 2,410 \log(1.6 \times 10^{-3} f[\text{Hz}] + 1) \quad (1)$$

各  $F_0$  変換音声における平均  $F_0$  及び平均ピッチを Table 4 に示す。

次に,  $F_0$  変換音声に対する心理実験の結果を示す。基準音声と比較音声の mel scale 上での平均ピッチ差に対する, ピッチ感及び速度感の評定平均を Fig. 3 に示した。

更に被験者ごとに, ピッチ感と速度感について回帰係数と切片をそれぞれ算出し, それらの平均から全体

Table 4 実験3の  $F_0$  変換音声

	90%	100%	111.11%
男声2文章3 (1.751 s)	143.0 (34.0) 214.7 (46.3)	158.9 (37.7) 235.9 (50.4)	176.7 (41.9) 259.2 (54.8)
女声2文章3 (1.851 s)	232.1 (51.1) 328.7 (50.1)	255.5 (57.0) 356.5 (55.3)	284.0 (62.5) 389.7 (57.8)

上段：平均  $F_0$  [Hz] (S.D.)

下段：平均ピッチ [mel] (S.D.)

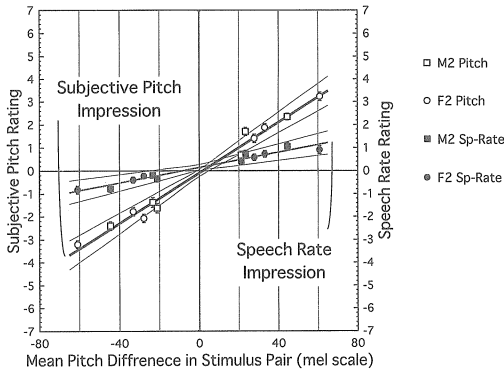


Fig. 3 STRAIGHT 分析合成による  $F_0$  変換音声のピッチ感と発話速度感  
基準音声と比較音声の平均ピッチ差ごとの評定平均と S.E.、及び回帰直線と 95%信頼区間。

での回帰式とその信頼区間を求め、図中に示した。

この回帰式により、刺激対の平均 mel 値の差分とピッチ感の評定値とを関連づけることができる。ここで、先の実験2とこの実験3の素材文章は共通であり、被験者も共通である。更に両実験とも STRAIGHT 分析合成によって変換した音声刺激を含んでいる。すると、双方の実験で観測されたピッチ感の評定値を等価なものと仮定して、両実験をつなぐ媒介変数と見なすことも可能である。従って、実験2で観測された話速変換に伴う主観的ピッチ感の変化量を基に、この実験3での回帰予測式を介して対応する mel 値を求めることができる。それにより、発話速度の操作に伴う、主観的ピッチ感の認知的バイアスの程度を定量的に扱い得ると考えられる。

なお、従来から評定尺度法で測定された音高に関する評定値は、音の高さを mel scale 上に変換した値と線形な関連があることが知られている<sup>19)</sup>。そこで、この実験で測定された音声のピッチ感についても、確認のために評定値を基準変数として回帰分析を行った。そして、説明変数に当たる変換音声の高さの指標として、 $F_0$  値の平均をそのまま用いた絶対周波数軸上のもの、その平均を対数変換した対数周波数軸上のもの、前述の分析で用いた mel scale 軸上のものの三つ

を比較したところ、mel 値の平均を用いたものの決定係数  $R^2$  が最も大きく、最も線形性が高かった。

### 2.4 話速変換が主観的なピッチ感に与えるバイアスに関する考察

ここまでの実験から、発話速度の低下に従って、ピッチ感が線形に低下しており、一定の認知的なバイアスが観測された。

ここでピッチ感の評定値を媒介変数としてとらえ、実験2と実験3で得られた回帰予測式を介することにより、認知的なバイアス量を推定した。結果としては、話速変換によって音声持続時間を150%に伸ばしてゆっくりにした場合、-14.01 mel 相当のピッチ感のバイアスが推定された。これは、男性2の音声データ (平均  $F_0$ : 158.88 Hz) では -10.42 Hz 相当のバイアスであり、女性2の音声データ (平均  $F_0$ : 255.49 Hz) では -11.71 Hz 相当のバイアスである。

本実験では、話速変換によって音声の持続時間が伸張された場合には、より低く認知されることが示された。くわえて、実験3で  $F_0$  のみを操作した音声では、 $F_0$  の上昇に伴って、話速の方もある程度速く認知されていた。この点を鑑みると、ピッチ感と発話速度感の間にはある程度の相関があり、言語音声における声の高さの認知は、そのような構造を持った認知的な枠組みの下でなされている可能性も指摘できよう。

### 3. 音声の発話速度の制御が話者の性格印象に与える影響

音声を通じて聞き手が認知する話者の性格印象は、話者自身の特徴を示す大切な情報である。実験4では、話速変換された音声における話者の性格印象を対象とする。性格特性5因子モデルに基づいた Big Five 尺度を用いて評価を行い、話速変換が話者の個人性情報に与える影響を検討する。

#### 3.1 実験4：発話速度の操作に伴う話者の性格印象の変化

話速変換された音声における話者の性格印象を対象とし、Big Five 尺度の形容詞チェック・リストを用いて評定尺度法で心理測定を行う。

### 3.1.1 実験4の被験者

実験1-3と同一の被験者が参加した。

### 3.1.2 実験4の音声刺激

素材の文章、話者とも実験2-3と同じで、実験2で使用された話速変換音声の中から、PICOLAによって、Fast 1 (80%), Original (100%), Slow 1 (125%), Slow 2 (156.25%) の4段階に変換されたものが使用された。

### 3.1.3 実験4の手続き

実験試行では、1 kHz 純音のピープ音に続いて、ターゲットとなる音声刺激が提示された。被験者は音声を聞き、話者の性格印象を評定した。

評定にあたっては、20項目で構成された形容詞チェック・リストが使用された。これは性格特性5因子モデル<sup>5)</sup>に基づいたBig Five尺度<sup>6)</sup>の項目を基に、各因子に安定して寄与する項目を精選した20項目からなる短縮版<sup>20)</sup>である。1因子につき4項目ずつで構成されている。

被験者は、各項目の形容詞の観点から話者の性格を評価し、「全くあてはまらない」から、「よくあてはまる」までの7段階(1…7)で評定した。1試行では、一つの音声刺激につき20項目の評定を行い、制限時間は60秒であった。全体では{(2話者×1文章)×4発話速度}の、計8試行を実施した。

また、刺激の提示順序についてはランダムな5系列

が作成され、被験者8名ごとに割り当てられた。更に、評定項目の評価用紙上での記載順序についてもランダムな4系列が作成され、先の音声刺激のランダム系列ごとに同数ずつ割り当てられた。

### 3.1.4 実験4の結果

話速変換音声の時間伸縮率に対する尺度合成得点の平均を、5種の性格特性ごとにFig.4に示す。各合成得点は、それぞれ4項目ずつの評定値から算出されたものである。

発話速度の操作に伴う性格特性の印象の変化をつかむため、特性ごとに平均合成得点から音声時間伸縮率への回帰曲線も示した。分析では、時間伸縮率を説明変数、伸縮率ごとの平均得点を基準変数として、1次式から3次式までの回帰式を順次あてはめた。自由度修正済み $R^2$ の値を指標とし、外向性でわずかに1次式のフィットがよかった以外は、他の4特性のすべてで2次式による回帰が最もフィットしていたので、全体として2次の回帰曲線を採用した。

図から読み取れるように、「温和な」といった評定項目に代表される協調性 (agreeableness) の性格印象では、速い発話で得点が最も低く、発話がゆっくりになるにつれて得点が上昇して、ややゆっくりな発話でピークとなり、更に遅くなると再び低下していた。一方、「計画性のある」という表現語に代表される勤勉性 (conscientiousness)、「独創的な」に代表され

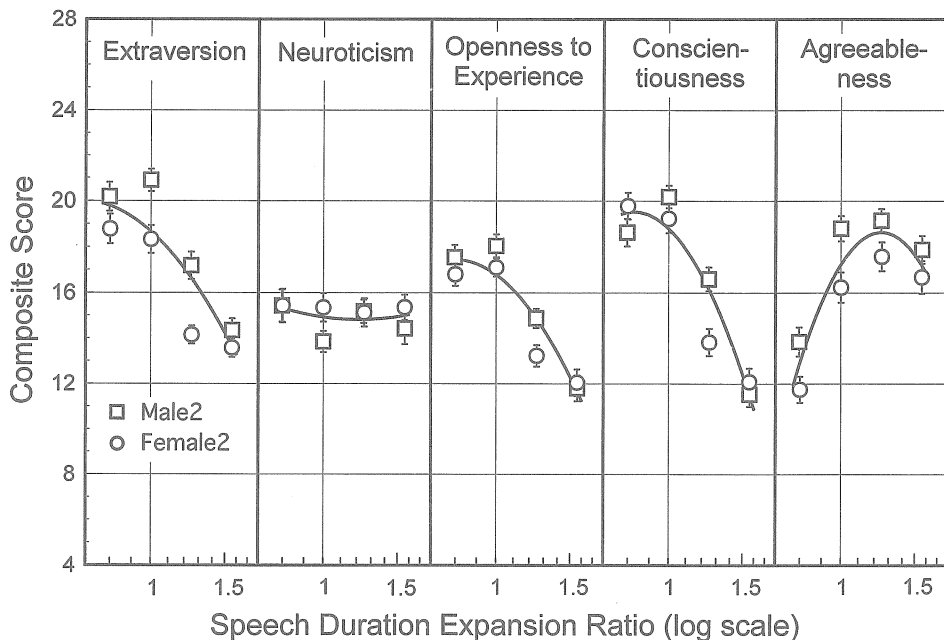


Fig.4 話速変換音声における話者の性格印象  
音声の時間伸縮率に対する性格特性ごとの合成得点平均とS.E., 及び回帰曲線。



る経験への開放 (openness to experience), 「陽気な」に代表される外向性 (extraversion) の各性格印象は、オリジナルの発話か、もしくはやや速い発話に得点のピークがあり、発話速度の低下に伴って得点が低下していた。なお、「悩みがち」に代表される情緒不安定性 (neuroticism) では大きな影響は見出されなかった。

次に各特性ごとでの分析結果を示す。各特性ごとに、その合成得点平均を変数として、被験者内 2 要因 [発話速度要因 (4 水準) × 話者要因 (2 水準)] の分散分析を行った。以下、特性ごとに整理する。

外向性 [評価項目: 話し好き, 無口な\* (逆転項目), 陽気な, 外向的] 発話速度要因が有意で ( $F_{(3,117)}=40.69, p<.0001$ ), 話速の低下に伴い、得点の低下が見出された。話者要因も有意で ( $F_{(1,39)}=26.95, p<.0001$ ), 男声サンプルが全般に高得点であった。また発話速度要因と話者要因の交互作用も有意で ( $F_{(3,117)}=3.13, p<.05$ ), 原音声のサンプルごとに話速変換の影響が現れる位置や強さに若干の違いがあるものと考えられる。

情緒不安定性 [悩みがち, 不安になり易い, 心配性, 気苦労の多い] 発話速度要因 ( $F_{(3,114)}=0.37, N.S.$ ), 話者要因 ( $F_{(1,38)}=2.02, N.S.$ ), 交互作用 ( $F_{(3,114)}=1.74, N.S.$ ) ともに有意差は見られず、話速変換による明確な影響は見られなかった。

経験への開放 [独創的な, 進歩的, 洞察力のある, 想像力に富んだ] 発話速度要因が有意で ( $F_{(3,117)}=38.70, p<.0001$ ), 主に話速低下に伴って得点が低下した。話者要因も有意で ( $F_{(1,39)}=13.86, p<.001$ ), 男声サンプルが高得点であった。更に交互作用も有意で ( $F_{(3,117)}=3.34, p<.05$ ), サンプルごとに話速変換による影響の現われ方に相違が見られた。

勤勉性 [いい加減な\*, ルーズな\*, 怠惰な\*, 計画性のある] 発話速度要因が有意で ( $F_{(3,117)}=65.92, p<.0001$ ), 主に話速低下に伴い、得点が低下した。話者要因は有意差なし ( $F_{(1,39)}=1.99, N.S.$ )。また交互作用が有意で ( $F_{(3,117)}=6.85, p<.001$ ), サンプルごとに影響の程度に相違が見られた。

協調性 [温和な, 寛大な, 親切的な, 協力的な] 発話速度要因が有意で ( $F_{(3,117)}=25.70, p<.0001$ ), 速い発話で得点が低く、発話速度の低下に伴い得点が上昇、後に再び低下した。また話者要因も有意で ( $F_{(1,39)}=29.13, p<.0001$ ), 男声サンプルが高得点であった。なお、交互作用には有意差は見られなかった ( $F_{(3,117)}=0.92, N.S.$ )。

この性格特性の測定が適当であったかどうかを検討するために因子分析を行った。1人の被験者が1試行

の中で評定した 20 項目の評定データを一つのオブザベーションと見なして、主成分分解による因子分析を行った。その結果、因子数はスクリー基準で五つであった。更にバリマックス回転を行い、各因子に最も高い因子負荷量を持つ項目を整理したところ、今回利用した Big Five 尺度短縮版での各性格特性を測定するための項目セットと同様の結果となった。従って、本実験のデータにおいても、上記の五つの性格特性はある程度独立に評価されており、その測定のための項目選択もほぼ妥当であったものと考えられる。

### 3.2 話速変換が話者の性格印象に与える影響に関する考察

話者の性格印象を個人性情報としてとらえた場合、話速変換はその情報に一定の影響を与えていることが見出された。認知される話者の性格印象は、性格特性ごとに、ある程度の独立性を持って構造的に影響を受けていた。すなわち、協調性に関しては、速い発話での評価は低く、発話速度の低下につれて上昇し、更に遅くなると再び低下した。それに対し、外向性、開放性、及び誠実性では、やや速い発話に評価のピークがあり、発話がゆっくりになるにつれて評価が低下した。なお、その影響を受ける程度は、発話者や音声自体の特徴によって若干異なることも示された。

音声の発話速度の 1 次元的な操作は、特定の性格特性のみに影響を与えるものではなかった。聞き手は話者の性格印象を、異なった特性因子ごとに、ある程度独立に、そして多元的に評価しているものと考えられる。一つ一つの性格特性の印象は、発話速度の変化に対して、比較的単純な関数関係で変化しているようである。しかし、それは音量の増加に伴って大きさの印象も大きくなるといった単調増加的な関係ではなく、いずれかの箇所にはピーク値を持つ関数であると考えられる。そして、それぞれの関数は各性格特性に固有で、かつ独立であるために、発話速度を変化させた連続体上では、変化のパターンが多様になっているものと考えられる。

実際のデータに即して検討してみると、話者要因の主効果が有意になっている性格特性もあることから、性格印象は当然のことながら話者の属性や音声サンプルに依存する部分があると考えられる。そして、原音声には既に抑揚などの形で一定の性格印象の情報が内在しており、それが性格印象の所与の初期値となりながらも、話速変換によってその評価が構造的に変化していると考えられる。なお、各性格特性はある程度独立に評価されていると考えられるが、話速変換に伴う変化パターンについては、外向性、経験への解放、勤勉性の 3 因子は比較的同様の傾向を示している。

#### 4. 総合考察

話速変換した音声においては、その平均  $F_0$  には有意な差が見られなかった。それにも関わらず、発話速度の上昇に伴って、主観的なピッチはより高く認知されており、逆に話速低下時には低く認知されていた。そして、ピッチ感の評定値を媒介変数とした精度の範囲内では、音声持続時間を 150% 伸長して発話速度を低下させた場合、約 -14.0 mel 相当の認知的バイアスが推定された。

これより、話速変換技術は、音高に関する音響的指標の下ではほぼ完成の域にあるものの、主観的なピッチ感といった認知的な側面においては一定のバイアスが生じていることが示されたと考えられる。

次に、話速変換は話者の性格印象に関して、組織的に影響を与えていることが示された。すなわち、協調性の印象は、発話速度の低下に伴って高くなり、ややゆっくりな発話でピークを、更に遅くなると再び低下する。それに対し、外向性、開放性、及び誠実性は、やや速い発話に評価のピークがあり、発話がゆっくりになるにつれて低下する。一方、情緒不安定性では、それほどの変化は見られない。

音声の発話速度の操作は、特定の性格特性のみに影響を与えるのではなかった。聞き手は、発話速度の変化の中で、話者の性格印象を性格特性因子ごとに多元的に独立して評価していると考えられる。それは、話速変換による影響の現れ方が性格特性ごとに異なり、更に各特性ごとに評定値のピークの位置が異なっているといった結果などにも現れている。

話速変換音声を扱っていると、変換によって聴感上の総体的な印象が、複雑に変容するといった経験をすることが多い。本実験の結果をふまえて、そのような印象の生起の解釈を試みる。まず各々の性格特性の印象は、発話速度に対して比較的単純な関数関係で変化するものと仮定する。すると、発話速度を操作した場合、それぞれの性格特性値は各特性に固有の関数のもとで変化する。総体的な聴覚印象は、変化した特性値が組み合わさったのものであると考えれば、話速変換に伴う総体的な聴覚印象の複雑な変化も、構造的にモデル化していくことができよう。

以上のように、話速変換操作によって音声をゆっくりと変化させるといったことは、聞き手にとって単に内容が分かり易くなるほかにも、聴覚印象上、多元的な影響を生じせしめていると言えよう。

更に、話速変換に伴う、先の主観的ピッチ感の変化と話者の性格印象の変化とは、基本的に別の事象であり、異なった機序で生起している可能性を指摘できよ

う。理由としては、主観的ピッチ感は発話速度の操作に対して安定して線形に影響が観測されるのに対し、話者の性格印象は各性格特性ごとに影響の現れ方が異なっている点。更に性格特性ごとに評定のピーク位置が異なっており、話速操作に対して必ずしも単純増加や単純減少の変化を示していない点を上げることができる。

さて、放送補助機器の開発においては、画像と音声との同期性といった制限のもとで、話速変換のメリットを生かす工夫が求められる。そのため、話速変換の伸縮率を音声フレーズ中でダイナミックに変化させる方法なども開発されてきている。しかし、このような複雑な変換が聴覚印象にどのような影響を与えるかは不明であり、今後の検討が必要であろう。

今後、話速変換技術がもたらす数々のメリットを生かしつつ、話速変換に伴う性格印象の変化に対する補正なども視野に入れた、次世代の技術開発が望まれる。また、感性情報処理技術の観点から、本報告で見出された性格印象の変化などを積極的に利用していくような、新たな技術の開発も期待できよう。

#### 5. まとめ

1. 話速変換した音声に対しては、発話速度の低下に伴って主観的なピッチがより低く認知されており、認知的なバイアスが生じていた。

2. 聴取者は、話者の性格印象を特性ごとに多元的に評価しており、音声の話速変換は、その印象に組織的な影響を与えていた。協調性の評価は、発話速度の低下に従って高くなり、更に遅くなると再び低下した。外向性、開放性、及び誠実性は、やや速い発話に評価のピークがあり、発話速度の低下と共に低下した。一方、情緒不安定性には大きな影響は見られなかった。

#### 謝 辞

本研究の実施にあたり、名古屋大学工学研究科の板倉文忠教授には PICOLA の利用に關しまして、また、和歌山大学システム工学部の河原英紀教授には STRAIGHT の使用につきまして、それぞれご快諾をいただきました。また、NHK 放送技術研究所の今井篤研究員、清山信正研究員、都木徹研究員には、音声刺激の一部の作成にご協力をいただきました。ここに深く感謝いたします。

なお、本研究の一部は、平成 11 年度日本心理学会第 63 回大会、及び日本教育心理学会第 41 回総会で発表いたしました。

また、本研究は、文部省の科学研究補助金(奨励研究(A)課題番号 10710073)の援助を受けました。

## 文 献

- 1) 都木 徹, “放送における話速変換: 話者や音環境の多様性への対応,” 音響学会誌 54, 533-538 (1998).
- 2) 中村 章, 清山信正, 池沢 龍, 都木 徹, 宮坂栄一, “リアルタイム話速変換型受聴システム,” 音響学会誌 50, 509-520 (1994).
- 3) 榎本義人, 在塚俊行, 今村俊樹, 伊福部達, 松島純一, “難聴者による話速変換音声の評価,” 音講論集 2-4-8, 317-318 (1994. 3).
- 4) 森田直孝, 板倉文忠, “自己相関法による音声の時間軸での伸縮方式とその評価,” 信学技報 EA 86-5, 9-16 (1986).
- 5) 柏木繁男, 性格の評価と表現 (有斐閣, 東京, 1997), pp. 107-123.
- 6) 和田さゆり, “性格特性語を用いた Big Five 尺度の作成,” 心理学研究 67, 61-67 (1996).
- 7) 柏木繁男, 和田さゆり, “5 因子モデル (FFM) による性格テストの依存的妥当性の検討,” 心理学研究 67, 300-307 (1996).
- 8) 小林哲則, 板橋秀一, 速水 悟, 竹沢寿幸, “日本音響学会研究用連続音声データベース,” 音響学会誌 48, 888-893 (1992).
- 9) D. Malar, “Time-domain algorithms for harmonic bandwidth reduction and time scaling of speech signals,” IEEE Trans. Acoust. Speech Signal Process. ASSP-27, 121-133 (1979).
- 10) D. Malar, R.E. Crochiere and R.V. Cox, “Performance of transform and subband coding systems combined with harmonic scaling of speech,” IEEE Trans. Acoust. Speech Signal Process. ASSP-29, 273-283 (1981).
- 11) 河原英紀, “聴覚の情景分析が生み出した高品質 VOCODER: STRAIGHT,” 音響学会誌 54, 521-526 (1998).
- 12) 河原英紀, 増田郁代, “時間周波数領域での補間を用いた音声の変換について,” 信学技報 EA 96-28 (1996).
- 13) 河原英紀, Alain de Cheveigné, “原理的に抽出誤りの存在しないピッチ抽出方法とその評価について,” 信学技報 SP 96-96, 9-18 (1997).
- 14) 都木 徹, 清山信正, 宮坂栄一, “複数の窓幅から得られた自己相関関数を用いる音声基本周期抽出法,” 信学会誌 J80A, 1341-1350 (1997).
- 15) 久保理恵子, 山田玲子, 河原英紀, “STRAIGHT 分析合成音声を用いた /r/-/l/ 音聴取訓練,” 音講論集 1-8-22, 383-384 (1998. 3).
- 16) S.S. Stevens and J. Volkman, “The relation of pitch to frequency: a revised scale,” Am. J. Psychol. 53, 329-353 (1940).
- 17) P.H. Lindsay and D.A. Norman, *Human Information Processing*, 2nd ed. (Academic Press, New York, 1977), p. 163.
- 18) 中川聖一, 鹿野清宏, 東倉洋一, 甘利俊一 (監修), 音声・聴覚と神経回路網モデル (オーム社, 東京, 1990), pp. 122-123.
- 19) 電子情報通信学会 (編), 新版 聴覚と音声 (電子情報通信学会, 東京, 1980), pp. 105-106.
- 20) 萩生田伸子, 繁樹算男, “因子数選択のための新基準の提案,” 心理学会第 59 回発表論文集, 数理・統計 1-66, 439 (1995).



内田 照久

昭和 63 年名古屋大学教育学部教育心理学専攻卒業。平成 5 年同大学院博士課程 (教育心理学) 修了。平成 6 年より大学入試センター研究開発部に勤務。主に外国語リスニング・テストの開発及び実施に関わる基礎研究の一環として、音声認知の研究に従事、現在に至る。博士 (教育心理学)。音響学会, 心理学会, 教育心理学会, 行動計量学会, 音声学会, 音楽知覚認知学会, The Acoustical Society of America 各会員。