

平均化された音声に対する好ましきについて

102G801 河本 達也

問 題

人はそれぞれ個性を持っている、人はそれによって、他者が誰であるかを識別することができる。例えば、顔であれば、目や耳、鼻や口等のパーツの形・位置の個体差を手がかりとし、複数の人物の中から特定の人物を識別することができる。音声においてもこれと同様にアクセントや声の高さ・速さ・声質等に個体差が大きく現れる。最近の研究では、モーフィング (morphing 変態) という技術を使って、視覚パターンである顔や、聴覚情報である音声を、個人間で合成することができるようになってきた。

このモーフィングをという技術を使って多くの人の顔を合成した顔パターンは平均顔と呼ばれ、Langlois & Roggman (1990) らの研究では、平均顔はありふれた特徴のない顔であるにも関わらず、非常に魅力が高く評価されることがわかっている。

河原 (1998) によって開発された高品質音声分析変換合成法 STRAIGHT 及びそれを用いたモーフィング技術を用いれば技術的には複数の個人の平均化された音声 (以後、平均音声) の作成は可能である。では、音声についても、視覚情報である顔のように平均化することで魅力が向上するのだろうか。

内田 (2002) は、男女各 2 名の音声再生速度を操作し、その音声参加者に聞かせて、話者の性格印象を評定させた。その結果、たとえば、話速の要因が有意で、普通速度 (オリジナル)、もしくは話速がやや速い発話速度の際、外向的で洞察力があり、勤勉であるという判断がなされる割合が高まることがわかっている。このような結果を考慮すれば、あらゆる特徴が平均化された平均音声必ずしも良い印象を与えるとは限らないと考えられる。

そこで本研究では、音声を平均化することで、声の魅力がどのように変化するかを確かめたい。

方 法

実験参加者 大学生 49 名 (男性 23 名, 女性 26 名) が実験に参加した。

装置 刺激としての音声を作成するために、IC レコーダーおよびマイクを刺激の提示及び反応取得の為、パーソナルコンピューター及びディスプレイを用いる。

刺激 男女各 8 名に協力してもらい、普段と同様に「こんにちは」と発声してもらい音声を録音した。そ

の音声データそれぞれを Audacity を用いて、音声「こんにちは」の音韻である「k」「o」「n」「i」「t」「i」「w」「a」それぞれの発声の再生開始時間を調べ、-10dB で正規化し STRAIGHT で使用可能な形式に変更した。

次に音声を STRAIGHT で分析し、スペクトログラムから特徴点を取ったこのとき、低い周波数の部分では特徴点がまとまっており、STRAIGHT スペクトログラムでは分かりにくいので、FFT analyser で先ほど調べた音韻の再生開始時間を基に、波形から周波数を調べ合成した。合成の順序については、男女それぞれ、2名の平均音声を 4 種作成し、その合成による 4名の平均音声を 2 種、最終的に 8 名による 1 つの平均音声に合成しこの際にできた合成音声 (男女各 7 種) とオリジナルの音声 (男女各 8 種) の計 30 種の音声を刺激として用いた。

手続き 実験参加者一人につき一台のパソコンとヘッドホンを用意し、音声がなることを確認してもらった。実験開始前に音声が何度でも確認できることと、あまり深く考えずに回答してもらうように指示を行った。その後実験用プログラムを起動し、男女各 15 の音声の印象について各自の速度で回答してもらった。提示される音声の順序は実験参加者毎にランダムに提示された。

結 果

得られた声に対する印象得点について、被験者の性差 (2 水準: 男性, 女性) × 声の性 (2 水準: 男性, 女性) × 音声の合成レベル (4 水準: 1 名, 2 名, 4 名, 8 名) の 3 要因の分散分析 (被験者の性: 被験者間要因・声の性: 被験者内要因・音声の合成レベル: 被験者内要因) を行った。以下に統計的に有意であった点をもとに結果の主要な特徴を示す。

音声の自然さ 合成レベルが上がるにつれて、男女の音声共に自然さの得点が減少したことがわかった (図 1)。

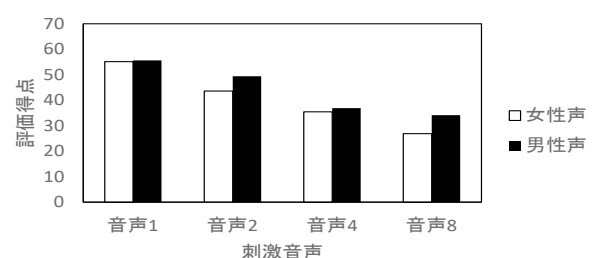


図 1 実験結果 (音声の自然さ)

音声の親近性 合成した音声の人数が増えるほど聞きなれた音声とは聞こえなくなっていく (図2)。

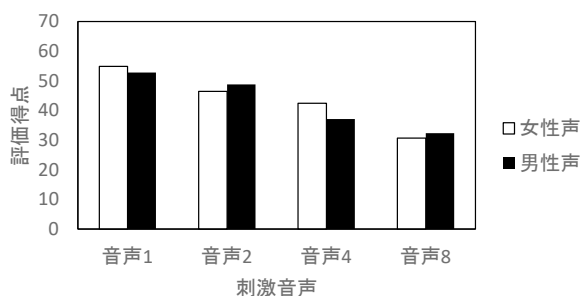


図2 実験結果 (音声の親近性)

魅力性 男性声よりも女性声の方が総じて魅力的だと評価され、合成レベルが進むにつれて魅力が低下することがわかった (図3)。

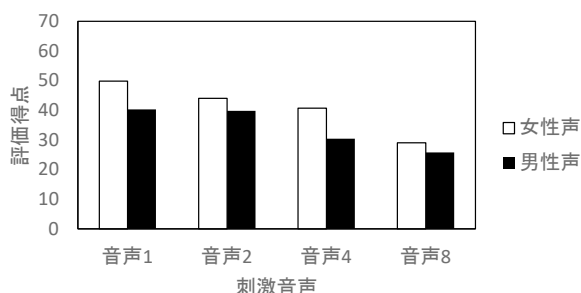


図3 実験結果 (音声の魅力性)

外向性 女性声の方が男性声よりも外向的だと評価され、合成が進むにつれて外向的だと評価されなくなることがわかった (図4)。

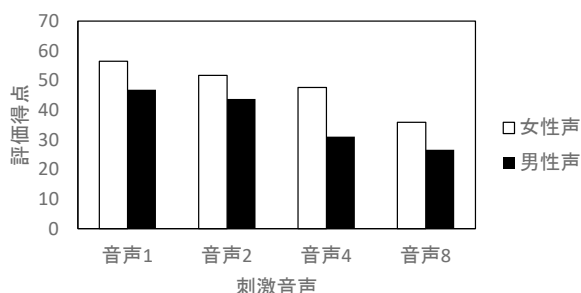


図4 実験結果 (音声の外向性)

情緒安定性 女性声の方が男性声よりも情緒的に安定していると評価され、合成が進むにつれて情緒が安定していないように聞こえたことがわかった (図5)。

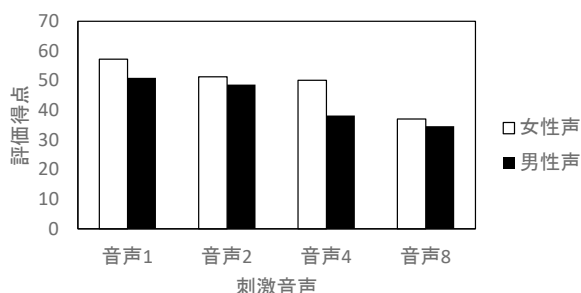


図5 実験結果 (音声の情緒安定性)

勤勉性 合成人数が8名になると、1~4名のときに対して、勤勉でないと評価される傾向にあることが

わかった (図6)。

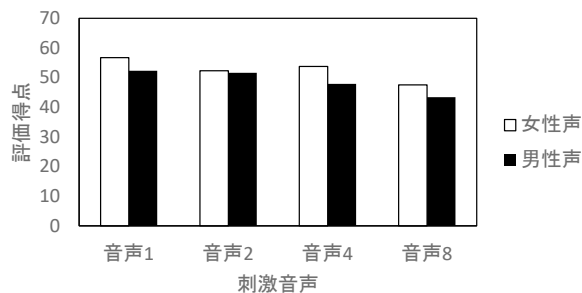


図6 実験結果 (音声の勤勉性)

協調性 女性声の方が男性声よりも協調的と評価され、合成が進むにつれてその評価は低下した (図7)。

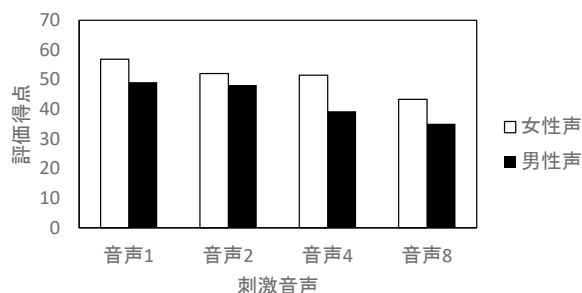


図7 実験結果 (音声の協調性)

考 察

全ての評価項目において、合成レベルに有意差がみられ、合成レベルが上がるにつれて各項目得点が減少する傾向があったことから、音声を合成し、音声を平均化するにつれて、平均顔と同様に評価が良くなるわけではないことが分かった。また、魅力・外交性・情緒安定性・協調性得点において声の性に有意差がみられ、女性の声の方が男性の声と比べ該当項目の印象得点が良く評価されることが分かった。しかしながら音声の自然さ得点において合成レベルによる有意差がみられたため、音声の不自然さが実験参加者の印象評価に大きく影響を与えている可能性があると考えられる。そのため、合成レベルが上がるにつれて音声の不自然さが現れないように音声合成を行う技術が必要になると思われる。

引用文献

河原英紀 (1998). 聴覚の情景分析が生み出した高品質 VOCODER: STRAIGHT 日本音響学会誌, 54, 521-526.
 Langlois & Roggman (1990). Attractive faces are only average. *Psychological Science*, 1, 115-121.
 内田輝久 (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.