

個人の同定における声質としゃべり方の優位性

062G013 沖田 唯・062G030 田室 恵

問題

これまでの研究で、非言語的コミュニケーションに関しては視覚的なものの方が多くとりあげられており、聴覚的な研究は少なかった。それは、技術的な問題で聴覚的コミュニケーションの研究がやりにくかったことも一因であり、それだけ研究例は少ない。

しかし、河原(1998)らによって STRAIGHT が提案された。STRAIGHT は、音声知覚の研究用に開発されたもので、音声の性質やピッチを操作するためのツールである。これによって可能になった技術に音声モーフィングがある(河原, 1998)。音声モーフィングとは、ある特徴をもつ音声から別の音声へ中間の音声を作成し、徐々に変化していく様子を表現することができる技術であるが、STRAIGHT を用いれば、音声を複数のパラメータに分割して独立に操作し、合成することが可能になる。

音声の個性は、音声器官の違いとしゃべり方に分かれる。音声器官の違いは先天的なものであり、しゃべり方の違いは後天的なものである。後天的なものはある程度まねることができるが、先天的な特徴は、変えることができない。この特徴は話者の識別や同定に有用な手がかりとなる(日本音響学会, 1996)。

実験 1

目的

本研究では、声質としゃべり方がどの程度個人の同定に影響しているのか、STRAIGHT の音声モーフィング法を用いて研究を行う。

方法

実験参加者 大学生 12 名が実験に参加した。

刺激 20 代女性 A さん、B さん 2 名による音声「こんにちは」を採集し、①A さんによる通常の発声(標準語風)と、②B さんによる訛りの入った発声(関西弁風)の間で声質としゃべり方をそれぞれ 10% 刻みで独立に操作してモーフィング合成を行い、121 の刺激音声を作成した。

手続き 作成した 121 個の刺激音声それぞれについて、A さん、B さん、刺激音声の 3 音声を 500ms 間隔で提示し、参加者には刺激が A さん B さんのどちらにより近く聞こえるかを強制選択で反応してもらった。

また、A さん、B さんの提示順序は参加者間でカウンターバランスを取った。なお、刺激音声と実験プログラムを作成し、動作などの確認を行うため、事前に筆者らによる予備実験を行った。

結果

実験の結果として得られた人物の判断率を図 1 に示した。

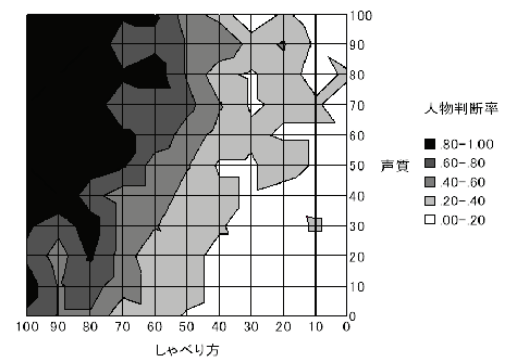


図 1 人物の判断率

横軸はしゃべり方についてのモーフィング率を、縦軸は声質のモーフィング率を、B さんが含まれる割合で示している。色分けされた領域は 12 名の参加者によって B さんと判断された割合を示している。

左上に原点をもつ領域は、80%以上の確率で B さんと判断された部分である。この部分は、横軸(しゃべり方)よりも縦軸(声質)方向に大きく広がっている。右下に原点をもつ領域は 80%以上の確率で A さんと判断された部分であるが、これも縦軸方向により大きく広がっている。このことは、声質がモーフィングによって他者に置き換わっても(縦軸の変化)、比較的安定して原点の人物として認識されていたことを示すものと言えよう。その一方で、しゃべり方が他者に置き換わること(横軸の変化)は、判断率に大きく影響している。判断率 50%を含む領域も縦軸に沿って広がっており、このことも、しゃべり方のわずかな変化が人物判断を A さんから B さんに分ける境界となったことを示している。

考察

本実験の結果は、話者の特定において、声質よりもしゃべり方の情報が大きく用いられることを示すものであった。しかしながら、声質は人それぞれで異なることから、それが話者の特定に有効でないということ

は意外に感じられた。また、実験前に行った予備実験の結果と整合していなかった。図2は、筆者らを含む3名が行った音声識別実験の予備実験結果である。

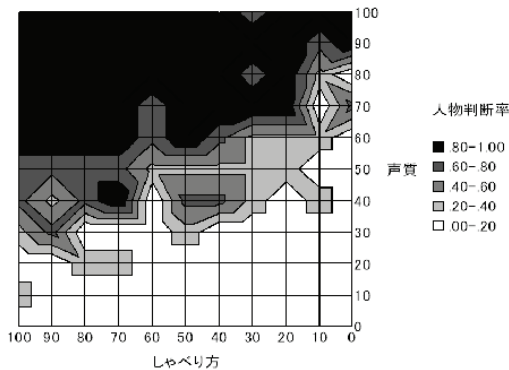


図2 予備実験での人物判断率

データ数は少ないが、この図を見る限り、本実験の結果と異なり、声質がしゃべり方よりも話者同定に寄与しているように思われる。なぜこのような結果の違いが生じたかについて、推測されたのは、予備実験では実験参加者が（話者の音声の録音作業等を行った関係で）話者と面識があったのに対し、本実験の参加者は話者を知らなかったという点である。話者とコミュニケーションを行う経験が、声質についての記憶を形成し、それが話者同定に用いられるのかもしれない。

そこで、実験2では、筆者ら2人の声を使い、同様な音声モーフィングにより刺激を作成した上で、筆者らと親しい友人を実験参加者として、音声識別実験を行った。

実験2

目的

実験1の結果および予備実験の結果から、話者とのコミュニケーション経験が、話者の同定に影響を与える可能性が示唆された。そこで実験2では、筆者ら2名の声を用いて、実験1と同じ方法で刺激を作成し、実験参加者には筆者とある程度以上コミュニケーション経験のある人物に協力してもらった。

方法

実験参加者 筆者らと親しい友人12名が参加した。

刺激 筆者ら2名(以下筆者C, 筆者Dとする)の、標準語風、関西弁風のそれぞれ2種類の「こんにちは」という音声を採用した。その音声を STRAIGHT で分析し、実験1と同様に、声質としゃべり方をそれぞれ10%刻みで独立に操作してモーフィング合成を行い、121の刺激音声を作成した。

手続き 作成した121の刺激音声それぞれについて、筆者C, 筆者D, 刺激音声の3音声を500ms間隔で

提示し、参加者には刺激が筆者C, 筆者Dのどちらにより近く聞こえるかを強制選択で判断してもらった。

結果

実験の結果として得られた人物の判断率を図3に示した。

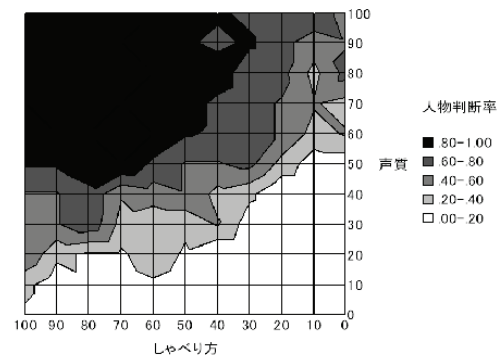


図3 実験2での人物判断率

図の見方は実験1と同様で、図の横軸はしゃべり方についてのモーフィング率を、縦軸は声質のモーフィング率を、筆者Dが含まれる割合で示している。左上に原点をもつ領域は、80%以上の確率で筆者Dと判断された部分である。この部分は、実験1と比べると縦軸(声質)の広がり小さくなり、横軸(しゃべり方)方向への広がりが大きくなっている。右下方向に原点をもつ領域は80%以上の確率で筆者Cと判断された部分であるが、こちらも同様に、横軸方向により大きく広がっている。このことは、しゃべり方がモーフィングによって他者に置き換わっても(横軸の変化)、比較的安定して原点の人物として認識されていたことを示すものと言えよう。その一方で、声質が他者に置き換わること(縦軸の変化)は、判断率に大きく影響している。判断率50%を含む領域も横軸に沿って広がっており、このことも、声質が人物判断を筆者Cから筆者Dに分ける境界となったことを示している。

考察

実験2の結果は、話者の特定において、しゃべり方よりも声質の情報が大きく用いられることを示すものであった。このように本研究の結果、話者とコミュニケーションを行う経験が、声質についての記憶を形成し、それが話者同定に用いられることが示唆された。

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.
- 日本音響学会(編)(1996). 音のなんでも小事典 講談社