

聴取による話者識別における鼻音・口音の非対称性と音響的距離との関連*

◎網野加苗, 菅原勉 (上智大・外), 荒井隆行 (上智大・理工)

1. はじめに

音声には発話の意味内容である言語学的情報以外に, 話者の感情や話者個人に関する情報なども含まれている [1]. 人間が音声のみを聞いて話者を識別する際, 聴取する音の種類によって識別のしやすさは異なり, 自動話者識別の場合と同様, 言語の種類を問わず, 母音や鼻音などの有声自鳴音を用いると有効であることが分かっている [2-4].

聴覚的な有声自鳴音の識別のしやすさには, 何らかの音響的な根拠があると考えられる [5]. 本研究では, まず聴取による話者識別実験によって識別のしやすい音の種類を確認し, 話者の識別のしやすさとその音響的対応について検討する.

2. 話者識別実験

実験条件は表 1 の通りである. 本実験の方法及び実験文は基本的には先行研究 [4] に基づいているが, 本実験では話者数が 10 名と多かったため, 聴取者の負担も考慮して刺激語の子音部は歯茎付近で調音される 6 種類と 3 種類の鼻音に絞った.

表 1. 実験条件

話者	男性 10 名
聴取者	話者全員をよく知っている男性 5 名
実験文	「～党を支持します」というキャリア文の「～」の部分に /aCaCaCa/ という /a/ と子音 C から成る構造を入れて読ませた
刺激語	実験文の /aCaCaCa/ から抜粋した最後の /Ca/. ただし C に入る子音は /l/d/s/z/h/fj/m/n/
録音	SONY TCD-D8 を用いて防音室で DAT に録音 (48kHz, 16bits)

聴取実験は一名ずつ防音室で行った. 9 種類の刺激語を 1 名の話者につき 5 標本ずつランダムに提示し, 推測した話者を回答用紙に記入させた. 各刺激語における話者識別の正答率を表 2 に示す. 聴取させた刺激語の種類別の繰り返し回数は, それぞれ 250 回

表 2. 刺激語別の識別正答率

刺激語	正答数 (250)	正答率 (%)
/na/	215	86.0
/nja/	214	85.6
/ma/ /za/	202	80.8
/sa/	197	78.8
/ja/	196	78.4
/da/	195	78.0
/ra/	186	74.4
/ta/	184	73.6

(話者 10 名 × 5 標本 × 聴取者 5 名) となっている.

表 2 からは先行研究 [4] における結果と同様, 鼻音が上位を占めていることが分かる. 口音に関しては調音方法などによる目立った傾向は見られないが, /ta/-/da/, /sa/-/za/ の対において, 無声音より有声音で正答率が高くなっている.

閉鎖音-摩擦音, 破裂音-摩擦音, 無声音-有声音などの 2 群間で *t* 検定を行ったが, 有意な差はみられなかった. 口音と鼻音の 2 群に関しても同様に 5% 水準では有意な差が見られなかった ($p = 0.057$).

3. 音響分析

上述の聴取結果の違いを音響的に説明することを目的として, 刺激語をスペクトル距離によって評価した. 刺激語は全て /Ca/ という同じ構造から成っているため, 聴取実験の結果の違いは主に子音部あるいは子音部から母音部への遷移区間によるものであると考えられる. そこで今回は実験の刺激語のうち, 比較的音響分析がしやすい /ta, da, sa, za, ma, na/ の 6 種類の子音部を分析対象とした. 調音方法別に表 3 に示すような基準を設け, Praat 4.2.1.7 を用いて子音部を抜粋した. なお, その際のフレーム長は 30ms とした.

抜粋後, 話者 10 名全員の全 5 回分の発話に関して

* The relationship between the asymmetry of the nasal and the oral sounds in human speaker identification and their acoustic distances. By Kanae Amino, Tsutomu Sugawara, and Takayuki Arai (Sophia University).

総当たり方式(50×50)でケプストラム距離を求めた。さらに、各話者について6種類の子音別にケプストラム距離の話者内、話者間での平均値を求め、話者内の値に対する話者間の値の比を算出した。比の値が大きいくほど話者による違いが大きい、つまり、より個人性を反映している可能性を意味している。子音別の話者内及び話者間の距離と、同じく子音別の、各話者及び全員の平均値をそれぞれ図1, 2に示す。

表3. 抜粋基準

/n/, /d/	解放を含む、母音への移行前の区間
/s/, /z/	母音への移行前の摩擦定常部
/m/, /v/	母音への遷移前の鼻音マーマー区間

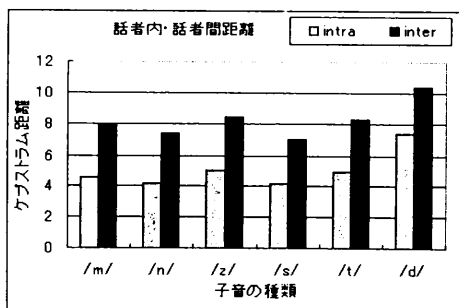


図1. 子音別 話者内・話者間距離

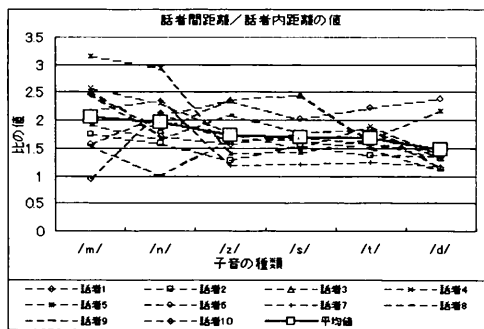


図2. 子音別 話者間距離/話者内距離の比の値

図2から話者内・話者間距離の比の値は鼻音 /m/, /n/ で大きく、閉鎖音 /t/, /d/ で小さくなっていることが分かる。この比の値に関して一元配置の分散分析を行ったが、子音による差は有意ではなかった ($p = 0.075$)。ただし、鼻音/口音の2群間でt検定を行ったところ、有意な差が見られた ($p = 0.0038$)。

4. 考察

聴取による話者識別実験の正答率とケプストラム距離の対応を見るために、表4において両者における子音の順位を示す。統計的には識別正答率とケプストラム距離の間に有意な相関は見られなかったが、調音方法別に鼻音/閉鎖音/摩擦音という順に話者の個人性をより多く反映している可能性があるという点において両者で共通の傾向が見られる。

表4. 識別正答率と音響的距離の対応

順位	識別正答率	音響的距離
1	/n/	/m/
2	/m/	/n/
3	/z/	/z/
4	/s/	/s/
5	/d/	/t/
6	/t/	/d/

5. おわりに

音声の聴取によって話者を識別する際に有効であるとされる音がなぜ有効なのかを音響的に説明することを目的として、聴取実験と音響分析を行った。聴取実験の結果、鼻音が含まれる音声構造を聞いた場合に、口音が含まれる構造を用いた場合よりも話者の識別がしやすいという傾向が見られた。実験の刺激音のうち、6種類の子音部に関してケプストラム距離を求めたところ、口音よりも鼻音の方が有意に話者間の距離が大きいという結果が出た。

その理由としては、鼻音の調音において口音よりも多くの生理学的特徴が関与することが考えられる。鼻腔・鼻咽腔・副鼻腔などは個人によって形状が異なるため、その共鳴特性は個人性を反映しやすいものと考えられる。話者の生理学的な個人性に関する他の研究 [6] にもある通り、今後は生理学的観点からも鼻音の特性に注目していく必要があるだろう。

参考文献

- [1] 新美(著), 坂井(編), 『音声認識』, 共立出版 東京, 1979.
- [2] 松井, 古井, I. Pollack, “連続音声の中音節による個人性知覚” 音講論集秋, 379-380, 1993.
- [3] Sambur, M.R., “Selection of acoustic features for speaker identification,” *IEEE Trans. ASSP*, vol. 23 (2), 176-182, 1975.
- [4] 網野, “聴覚による話者識別における日本語音韻の特性” 信学技報, SP2004-37, 49-54, 2004.
- [5] Fant, G., *Acoustic Theory of Speech Production*, The Hague: Mouton, 1960.
- [6] 北村, 本多, “母音発声時の声道形状における不変部位について,” 日本音声学会講演論文集, 105-110, 2003.