

# 雑音下における話者識別に発話内容が及ぼす影響について\*

◎網野加苗, 荒井隆行 (上智大・理工)

## 1 はじめに

聴取による話者認識の特性を調べることは、法工学などにとって有用である。

話者認識の聴取実験では、呈示する刺激音の種類によって正答率が異なることが報告されている [1]. 著者らによる今までの研究では、鼻子音と母音から成る音節が、口子音と母音から成る音節よりも有効であることが示された [2-4]. また、音節に含まれる子音の調音方法別に見ると、「鼻音-摩擦音-閉鎖音」のように、響きの程度 (sonority scale [5]) 順に並び、調音点別には唇音よりも歯茎音で正答率が高くなることが示された [2-4].

一方、雑音下の単音節認識では、鼻子音の調音点の弁別が難しいことや後続母音へのフォルマント遷移による影響があることが指摘されており [6], 聴取による話者認識における鼻音の有効性が保たれる保証はない。雑音下における話者認識について調べた研究はそれほど多くないが、クリーン音声による話者認識よりも基本周波数の影響が大きくなることなどが指摘されている [7]. しかし同研究では刺激音として文を呈示しており、その発話内容の影響までは考察されていなかった。

本研究では、雑音下で聴取によって話者を同定する際に、発話内容による正答率の違いがどのような影響を受けるのかを調べるべく、実験を行った。

## 2 聴取実験

### 2.1 音声資料

本実験に用いた音声資料の詳細を表 1 に示す。電子協日本語共通音声データから、比較的静かな環境で録音された 4 名の関東方言話者の音声を選択して用いた。発話内容は 9 種類の単音節で、いずれも子音に後続する母音は /a/ である。これは、実験を単純にするためであり、今までの実験結果との比較を容易にするためでもある。また、調音点のバリエーションの多い歯茎音を用いることで、発話内容の比較を行う。

雑音には、NOISE-X 雑音コーパスからバブル雑音と白色雑音を選択した。刺激作成前に、標本化周波数を音声コーパスに合わせて 48 kHz にリサンプルした。

### 2.2 刺激作成と手続き

MATLAB を用いて表 2 に示すような刺激音を作成した。雑音なしの音声以外に、バブル雑音、白色雑音をそれぞれ 3 種類の S/N で重畳した音声など、7 条件で作成した。図 1 に 2 種類の雑音の周波数特性を示す。

雑音重畳音声では、単音節の前後にそれぞれ 300 ms の空白を入れ、立ち上がり及び立ち下がり 200 ms 分の丸めをつけた雑音を重畳した。

Table 1 Speech and noise corpora used in this study

音声	コーパス	電子協日本語共通音声データ 社団法人日本電子工業振興協会
	標本化・量子化	48000 Hz, 16 bit
	使用データ	単音節の発話
	話者	男性 4 名, 関東方言話者
	使用した音節の種類	/da, ja, ma, na, nja, ra, sa, ta, za/
雑音	コーパス	NOISEX-92
	標本化・量子化	19980 Hz, 16 bit
	使用した雑音の種類	バブル雑音, 白色雑音

\* Effects of the stimulus contents on perceptual speaker identification under noise, by AMINO, Kanae and ARAI, Takayuki (Sophia University).

Table 2 Created stimuli

S/N	10 dB, 5 dB, 0 dB	
刺激音の種類	#1	雑音非重畳音声
	#2	バブル雑音重畳 (0 dB)
	#3	バブル雑音重畳 (5 dB)
	#4	バブル雑音重畳 (10 dB)
	#5	白色雑音重畳 (0 dB)
	#6	白色雑音重畳 (5 dB)
	#7	白色雑音重畳 (10 dB)

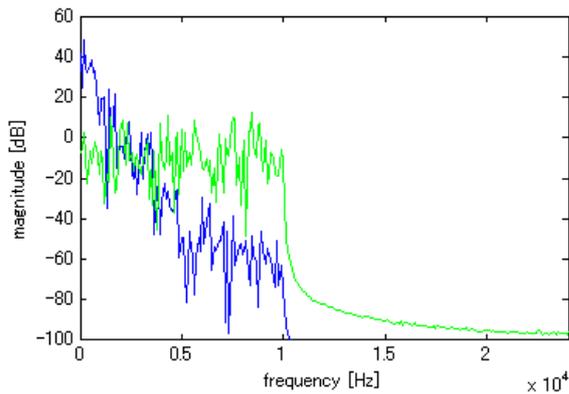


Fig.1 Frequency properties of the 2 noise types: babble noise (blue) and white noise (green).

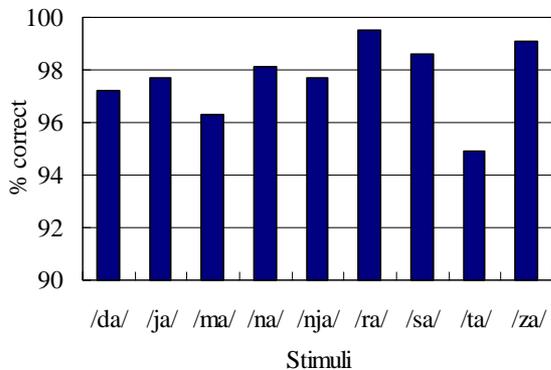


Fig.2 Speaker identification accuracy by speech without noise

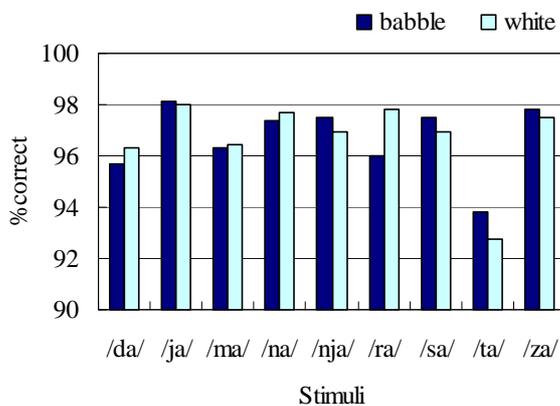


Fig. 3 Speaker identification accuracy under noise environments (according to the noise types)

刺激は、4名の話者から2名を選ぶ組合せで、可能な全ての組合せが含まれるようにランダムに並べ、AXB法によって呈示された。A及びBには、同じ単音節の雑音非重畳音声（#1とは異なるトークン）を呈示し、Xに7条件（#1～#7）の刺激音を呈示した。

聴取者は日本語を母語とする健聴者9名である。聴取実験は1名ずつ防音室で行った。実験課題はXの話者がAとBのどちらの話者と同一かを答えることである。刺激音は一回のみ呈示された。

### 2.3 仮説

実験に際し、以下の仮説を立てた。

1. 雑音下における話者認識でも、非雑音環境下と同様、刺激音の種類による正答率の違いがある。
2. 雑音の種類によって刺激音の有効性が異なる。
3. S/N が大きいほど、話者同定の正答率が上昇する。

## 3 結果

### 3.1 原音声による話者識別正答率

本研究では、課題の評価として、話者同定の正答率を用いた。#1の刺激音に対する話者同定の正答率を図2に示す。

仮説1に関して、いずれの刺激音間にも統計的に有意な差はなかったが、以下のような傾向が確認された。まず、先行する実験の結果と同様 [1-4]、子音部が有声音である /da/ /za/の方が、子音部がそれに対応する無声音である /ta/ /sa/ よりも正答率が高いという結果になった。さらに子音の種類別に見ると、今回の実験では /ra/ で最も正答率が高く、続いて子音部が摩擦音である /za/ /sa/、鼻

音・接近音の /na/ /nja/ /ja/, 最後に閉鎖音の /da/ /ta/ という順になった. 先行研究 [2-4] で見られた, 同じ鼻音でも歯茎音 /na/ /nja/ が両唇音 /ma/ より高正答率となるという傾向は, 今回の実験でも見られた.

### 3.2 雑音の種類の影響

図3は, #2~#7の刺激音の結果を雑音の種類別に示したものである. 全体的に正答率は高く, 雑音下の単音節でも十分に話者を同定できることが分かる.

統計的には雑音の主効果は確認されなかったが, これは今回の結果における天井効果が原因となっている可能性が高い. 個々の刺激音との関連について, 特に目立った傾向は観察されないことから, 仮説2は支持されなかった.

### 3.3 S/Nの影響

実験結果を S/N 別にまとめたものを図4に示す. S/Nの主効果は有意ではなかったが, 仮説3のような単純な傾向にはならないことが分かる.

ここで, 調音方法別に見た傾向は, 以下のようによまとめられる.

- 閉鎖音では S/N の増加に伴い, 正答率も高くなる.
- 摩擦音では雑音の種類によって傾向が異なり, バブル雑音下では S/N が 5 dB の時に最も正答率が低いのに対し, 白色雑音下では /sa/ は閉鎖音と同様の傾向, /za/ は S/N が 5 dB の際に正答率が高かった.
- 鼻音はバブル雑音下では S/N が 5 dB の時に正答率が高く, 白色雑音下では逆に低かった.

## 4 まとめと考察

本研究では雑音を重畳した様々な単音節刺激を用いて, 4名の話者を同定する実験を行った. その結果, 全ての条件下で正答率が高かったため, 単音節中の子音によるランキングに統計的な有意差は見られなかった.

特に, 先行研究と比較して, 雑音非重畳音声における鼻音の有効性は, 本実験の話者では示されなかった. これは, 話者によって識別や同定に有効な発話内容の種類は異なるという報告 [8] や, 話者による違いだけでなく, 話者セットの設定 (誰と比較されるか) によっても結果が変わるという報告 [9] によって説明できる.

ここで, 表3に示すように条件別に単音節の正答率の順位を見てみると, 色分けしてある通り, 調音方法別に大体並ぶことが分かる. 閉鎖音 (青色系) は全体的に下位に来ており, 摩擦音 (黄色系) 及び鼻音・接近音 (赤色系) のどちらかが上位に来ている. 今までの研究では [2-4] 音韻セットの順序は響きの程度の順序と一致すると考えていたが, 必ずしもそうならないことが示された.

以上から, 話者及び話者セットごとに有効な発話内容は異なるものの, 有効な種類の音は調音方法別に順位付けできると考えられる. つまり, ある話者セット内で一話者を特定するためには, 特定の音韻が有効であるというよりは, ある音韻のセット (phonological class) の調音上のパラメータ (摩擦成分や鼻腔共鳴の特性など) が話者間で異なっていることが重要であり, そのパラメータは話者セットによって違うということである.

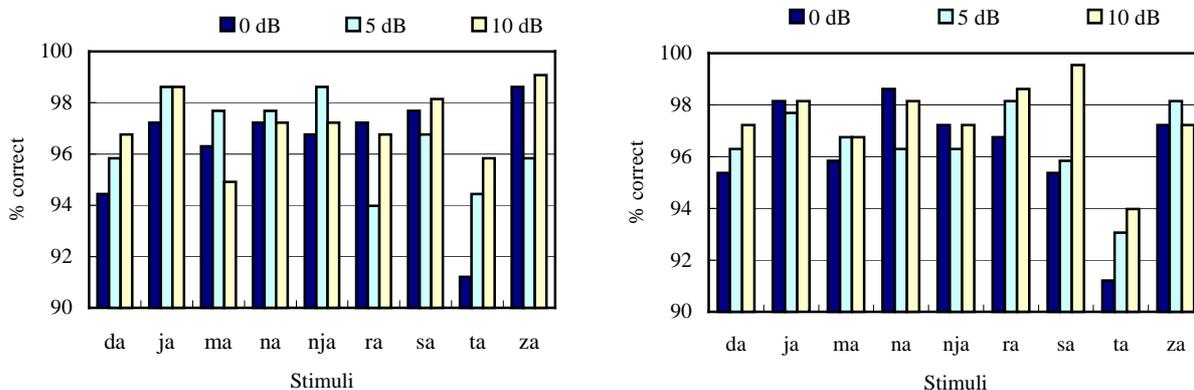


Fig.4 Speaker identification accuracy under noise environments (according to S/N): speech with babble noise (left) and with white noise (right)

Table 3 Rankings of the syllables under each condition

#1	#2	#3	#4	#5	#6	#7
ra	za	nja	za	na	ra	sa
za	sa	ja	ja	ja	za	ra
sa	ra	na	sa	nja	ja	na
na	na	ma	nja	za	ma	ja
ja	ja	sa	na	ra	da	za
nja	nja	za	ra	ma	na	nja
da	ma	da	da	da	nja	da
ma	da	ta	ta	sa	sa	ma
ta	ta	ra	ma	ta	ta	ta

また、仮説 3 が支持されなかった点に関して、本研究で使用した雑音は全て同一ファイルの同一区間を用いているため、S/N によって雑音の周波数特性に差が出ることは考えにくいですが、特にバブル雑音では S/N を変化させた際に、ある時点で特定の周波数成分が強調された可能性もある。よって、今後は時間軸に沿って周波数分析を行ってみる必要がある。

最後に、今回は雑音下での単音節による話者同定実験には前例がなかったため、課題の難易度の設定に問題があり、結果として天井効果が出てしまった。今後はより適切なデザインにより、検証実験を行う必要がある。

- [7] 橋本ら, 音響学会誌, 54(3), 169-178, 1998.
- [8] 松井ら, 音講論(秋), 379-380, 1993.
- [9] Bricker & Pruzansky, *Experimental Phonetics*, Lass (ed.), 295-326, Academic Press, 1976.

## 謝辞

本研究の一部は日本学術振興会特別研究員奨励費（17・6901）及び文部科学省私立大学学術研究化推進事業上智大学オープンリサーチセンター「人間情報科学研究プロジェクト」の助成を得た。

## 参考文献

- [1] 西尾, 言語生活, 158, 36-42, 1964.
- [2] Amino *et al.*, Proc. of Interspeech, 2025-2028, 2005.
- [3] Amino *et al.*, Acoust. Sci. Tech., 27(4), 233-235, 2006.
- [4] Amino *et al.* Acoust. Sci. Tech., 28(2), 128-130, 2007.
- [5] Selkirk, *Phonology and Syntax*, MIT Press, 1984.
- [6] Alwan *et al.*, Proc. ICPhS, 167-170, 1999.