

音声の個人性知覚における鼻音の寄与*

☆赤木隆之助, 北村達也 (甲南大)

1 はじめに

鼻腔は形状の個人差が大きく, 発話中も形状が変化しないため, 音声の個人性生成にとって重要な器官の1つであると考えられる [1]. 実際, 音声の個人性知覚に鼻腔の音響特性が寄与することが報告されている [2].

Amino ら [3] は音声において鼻腔から出力される成分を (1) 鼻音生成に伴って現れる成分, (2) 鼻音の前後に現れる成分, (3) 軟口蓋の振動に伴って現れる成分, (4) 声質に関連する成分, の4つに分類した. そして, 鼻音を含む文と含まない文の発話時に口腔と鼻腔から出力される成分を分離して収録し, これら4つの成分との関連を議論した. さらに, 男性話者が発話した鼻音を含む文と含まない文を対象に聴取実験によって話者間類似度を求め, 前者の方が類似度が高くなることを示した.

本研究では, Amino ら [3] が男性話者に対して行った実験を女性話者に対して実施する. そして, 音声において鼻腔から出力される成分や話者間類似度の判定における鼻腔の音響特性の影響に性差があるか否かを調査する.

2 口腔・鼻から出力される音声の分析

2.1 話者

話者は日本語を母語とする21歳または22歳の女性8名である.

2.2 方法

防音室にて, “ロンドンでミイラを見る”, “アラビアで油を売る”(以下“ロンドン文”, “アラビア文”と呼ぶ)の2文を2つの方法で3回ずつ録音した.

1つは通常モノラル録音である. マイクロフォン (SONY ECM-77B) とオーディオインターフェース (Roland Rubix24) を使用し, サンプリング周波数は48kHz, 量子化は16bitで録音した. もう1つは口腔と鼻腔から出力される音声 (以下 O-signal, N-signal と呼ぶ [3]) を分離できるマスク (Glottal Enterprises OroNasal Speech Tutor

Mask MA-1L) を使用した録音である. マイクロフォン (SONY ECM-77B) は2個使用した.

2.3 結果と考察

話者1名のロンドン文とアラビア文の O-signal と N-signal の音声波形とスペクトログラムをそれぞれ Fig. 1 と Fig. 2 に示す. 鼻音を含むロンドン文の N-signal では, 鼻音およびその前後にて振幅が大きくなっている. また, 鼻音を含まないアラビア文でも, ロンドン文よりは振幅が小さいながらも N-signal が観察される. これは Amino ら [3] と同様の結果である. この成分は, 軟口蓋の振動を通して鼻腔から出力されたものと考えられる [4]. つまり, 鼻音を含まない音声でも鼻腔の音響的影響を受けていることを意味している.

3 聴取実験

話者特性の知覚における鼻音の役割を調べるために話者類似性を評価する聴取実験を行った.

3.1 実験参加者

聴力に問題のない20代の男女 (男性15名, 女性5名) 20名が参加した. 参加者は話者と会話したことない.

3.2 刺激音

モノラル録音したアラビア文とロンドン文の音声の振幅を正規化し, 2話者を1対としたものを刺激音とした. 話者の組み合わせは36組 (異なる話者28組, 同一人物8組) で, 提示順を入れ換えたものを含む72対を用いた.

3.3 実験方法

実験は防音室にて行った. PC から出力された刺激音をヘッドフォンアンプ (Fostex HP-A3) と密閉型ヘッドフォン (Sennheiser HDA200) を使用して実験参加者の聴きやすいレベルで提示した. 実験前に5対の刺激音を聞かせ参加者自身が聴きやすい音量と実験の流れの確認を行った.

2人の話者の類似度を5段階 (1: 全く似ていない, 2: 少し異なる, 3: 少し似ている, 4: 似

*Contribution of nasal sounds in perception of speaker individualities. by AKAGI, Ryunosuke, KITAMURA, Tatsuya (Konan University)

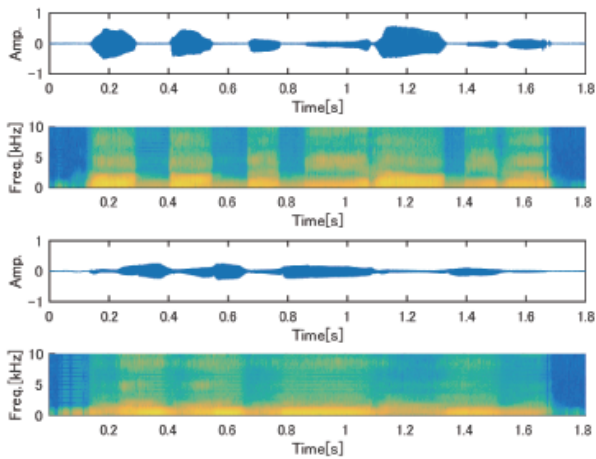


Fig. 1 Waveforms and spectrograms of O-signal (upper) and N-signal (lower) for the London-sentence.

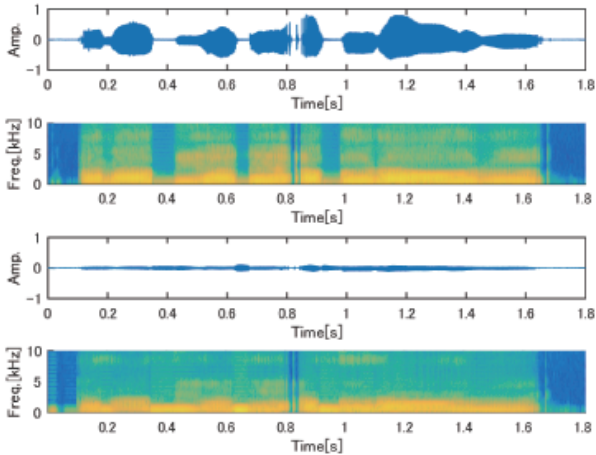


Fig. 2 Waveforms and spectrograms of O-signal (upper) and N-signal (lower) for the Arabia-sentence.

ている、5：同一人物）としてPCのスクリーン上で評価させた。1度だけ聴き直しを許可した。実験参加者には適宜休憩を取らせた。

3.4 結果と考察

ロンドン文とアラビア文の話者類似度マトリクスをそれぞれ Table 2 と Table 1 に示す。この値は実験参加者 20 人の平均値である。S1, ..., S8 はそれぞれの話者を表す。

2つのマトリクスを比べると、36 対中 21 対 (58.3%) で類似度が高くなっている。Amino ら [3] も同様の結果を報告している。特に S3-S4, S5-S8 の対でロンドン文の値が大きく上回っている。これらの結果は文に鼻音が含まれる場合にはそ

Table 1 Speaker similarity matrix for the London-sentence

	S1	S2	S3	S4	S5	S6	S7	S8
S1	5.00	2.43	2.00	2.53	1.55	1.53	3.58	2.30
S2		4.93	3.88	3.55	2.25	2.23	2.18	2.65
S3			4.95	3.95	2.08	1.88	2.45	2.45
S4				4.93	2.25	1.98	3.50	2.83
S5					4.95	2.43	1.68	3.25
S6						5.00	1.48	2.18
S7							4.95	2.40
S8								4.93

Table 2 Speaker similarity matrix for the Arabia-sentence

	S1	S2	S3	S4	S5	S6	S7	S8
S1	4.95	2.45	2.15	1.58	1.25	2.10	3.23	2.23
S2		4.95	3.70	2.93	1.98	2.88	2.33	2.93
S3			4.98	2.13	1.83	1.75	2.58	2.28
S4				4.93	1.65	1.75	2.43	2.35
S5					4.98	2.45	1.65	2.25
S6						4.95	1.80	2.73
S7							4.90	2.30
S8								4.98

の情報が個人性知覚に利用されることを意味している。

4 おわりに

本研究では、男性話者を対象に行われた Amino ら [3] の実験を女性話者を対象にして行った。そして、彼女らと同様の結果が得られた。これは個人性知覚における鼻腔の音響特性の寄与に性差がないことを示している。

謝辞 本研究の一部は 2018 年度科研費 (No. 17H02339) の支援を受けた。

参考文献

- [1] Dang *et al.*, JASA, 96(4), 2088-2100, 1994
- [2] Amino and Arai, ASJ, 30(2), 89-99, 2007
- [3] Amino *et al.*, Interspeech 2014, 1688-1692, 2014.
- [4] Dang *et al.*, JASA 139(1), 441-445, 2016.