

## 音素に着目した音声スペクトルの個人性の時期差の検討\*

◎ 網野加苗, 荒井隆行 (上智大・理工)

## 1 はじめに

現在までの著者らによる一連の研究 [1-7] では, 知覚による話者識別実験を行い, 表 1 に示すように音素によって話者の個人性への貢献度が異なることを示してきた. この違いを活用すれば, 自動話者認識などの性能を向上させることができるなど, その応用も期待される [8-9]. 音素に着目した話者認識手法の提案は特に近年増加しており, 日本語や英語において実用化に向けて研究が進められている [10-15].

一方, 話者認識やその他の音声技術に大きな影響を及ぼす要因の一つに音声の時期変動がある [16]. 音声の経時変化に関する研究は数多く存在し, 話者識別や話者照合, 音声認識, 音声データベースの構築において, 短期・長期に渡る音声変動の調査やその対処法が考えられてきた [17-26]. これらの先行研究における共通認識としては, 音声の経時変化は声道特性より音源特性で大きく, 声道特性に関しても 3 ヶ月程度で話者認識システムの性能を著しく劣化させてしまうことなどが知られている.

しかし, 音素ごとに個人性への貢献度が異なるように, 音素ごとに話者の個人性の経時劣化も異なる可能性がある. 本研究ではその調査の第一歩として, 話者識別における有効性が示された母音および鼻音のスペクトルを分析したので報告する.

表 1. 現在までの実験から得られた結果 [1-7]

聴取による話者識別に有効な音, 弁別素性	例
母音, 特に後舌母音	/a/ /o/ /u/ > /i/ /e/
[+nas], [CORONAL], [+voice] の弁別素性を持つ子音	/n/ /ɲ/
[PLACE] の素性は [+nas] 素性より優位にある	/n/ > /d/ > /m/ > /b/

表 2. 音声資料の録音

発話者	2名の日本語母語話者 (男性)
録音期間	2004年5月~2005年2月
時期差	1ヶ月ごとに計10回 (10ヶ月間)
録音場所	上智大学音声学研究室付属防音室
機材	マイク SONY ECM-MS957, レコーダ SONY TCD-D8
発話内容	単独発話の母音および鼻音, 4種類の文章, 自由会話など

## 2 音響分析

## 2.1 音声資料

2名の男性話者による文章の連続発話から抜粋された鼻音と単独で発話された母音および鼻音を分析対象とした. 音声資料の概要, 録音手順, 分析対象とした文章の全文を表 2, 図 1, 図 2 にそれぞれ示す. 録音は全て防音室内で 1 名ずつ行われた.

音声資料は DAT (Digital Audiotape) にサンプリング周波数 48 kHz, 量子化精度 16 bit で録音され, 分析前に 16 kHz にダウンサンプリングされた. 10 ヶ月間, 同内容の発話を毎月 1 回録音し, そのうち 1, 2, 3, 5, 7, 10 ヶ月目の音声資料を分析に用いた.

連続発話の録音では, 紙面上に書かれた文章を楽な声の高さおよび発話速度で朗読するよう教示した. 単独発話の母音および鼻音については, 指示された音を一定の声の高さで持続して発話するよう教示した.

2名の男性話者は録音時それぞれ 20 歳代と 30 歳代で, 長期海外滞在経験はなく, 健聴で, 歯列矯正や発声障害の既往もなかった. 共に両親の代からの日本語東京方言の母語話者であり, 両名とも音声学の訓練を受けたことがあった.

\* Investigation on speaker characteristics that appear on the spectra of each phoneme and their change over time, by AMINO, Kanae and ARAI, Takayuki (Sophia University).

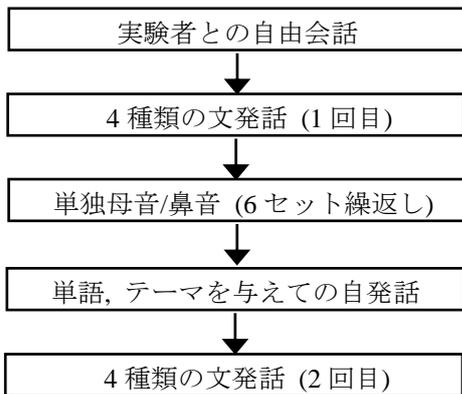


図 1. 録音の手続き

あるとき、北風と太陽が力比べをしま<sub>(m1)</sub>した。旅人の外套を脱がせた方が勝ちと決めて、ま<sub>(m2)</sub>ず北風から始めま<sub>(m3)</sub>した。北風は、なに、ひとま<sub>(m4)</sub>くりにしてみせようと激しく吹き立てま<sub>(m5)</sub>した。すると旅人は北風が吹けば吹くほど外套をしっかりと体にくっつけま<sub>(m6)</sub>した。今度は太陽の番になりま<sub>(m7)</sub>した。太陽は雲の間から優しい顔を出して暖かな光を送りました。すると旅人はだんだんよい心持ちになって、しまいには外套を脱ぎま<sub>(m8)</sub>した。そこで北風の負<sub>(m9)</sub>けとなりま<sub>(m10)</sub>した。

図 2. 連続発話「北風と太陽」の全文と/m/の抜粋箇所

## 2.2 分析条件

連続発話、単独発話ともに、定常的な部分を 30 ms ずつ抽出し、スペクトルを分析した。連続発話では当該の音素 (/m/) を目視によって見つけた後、手動により抜粋し、単独発話では持続した発話の中心部の 30 ms 間を自動的に抜粋した。

抜粋区間のフレームにハミング窓をかけ、512 点の FFT によって対数パワースペクトルを求めた。なお、最大値などによる正規化は行っていない。

## 3 結果

2 名の話者による連続発話および単独発話のスペクトルを以下、図に示す。全ての図に共通して、左の列が話者 1、右の列が話者 2 のスペクトルである。

上述の通り、初回録音を行った 5 月の音声データから、2 ヶ月目の 6 月、3 ヶ月目の 7 月、5 ヶ

月目の 9 月、7 ヶ月目の 11 月、最終回 10 ヶ月目の翌 2 月録音分の音声データを時系列に並べて示してある。連続発話は 2 回分、単独発話は 6 回分の繰り返し発話を重ね描きしてある。

### 3.1 連続発話からの抜粋音

発話文「北風と太陽」の連続発話から抜粋された /m/ のスペクトルを図 3 および図 4 に示す。紙面の都合上、図 2 に示した 10 個の抜粋箇所 (m1) から (m10) のうち、始めの 2 つの結果のみ掲載する。

図 3 と図 4 の比較から、同じ話者、同じ時期の同じ /m/ という音素の発話であっても、文脈による変動がかなり大きいことがわかる。1 回目と 2 回目の文発話の録音の間に、単独母音や自発話の録音をはさんでいるにも関わらず、同じ話者の繰り返しによる変動はさほど大きくない。しかし同じ話者の同じ文脈の /m/ であっても、時期を経ることによって特に 3 kHz 以上の帯域で不規則な変動が見られる。この 2 名の話者に関して言えば、話者間の違いは特に鼻音の第 1・第 2 フォルマンの振幅差に見られるが、この低次鼻音フォルマンにおける話者性は時間を経ても変化しない。また全体的に、FFT スペクトルよりもスペクトル包絡の方が分析に適していると思われる。

### 3.2 単独で持続発話された音声

単独発話の母音および鼻音 /m/ /n/ のスペクトルを図 5~8 に示す。図 5 の母音 /a/ では、第 2 フォルマンから第 3 フォルマンにかけてのスペクトルの谷および 7 kHz 以上のスペクトル形状が 2 名の話者間で大きく異なっており、この特徴は時期を経てもあまり変動しない。図 6 の母音 /i/ では、第 3 フォルマンより高い 3 kHz 以上の帯域で話者による違いが見られる。この違いも微細な変動を除けば、時間変化にロバストであると言える。

図 7 に示した単独発話の鼻音 /m/ のスペクトルでは、図 3 および図 4 の連続発話中の /m/ が文脈の影響を大きく受けていたことが改めて確認できる。経時的な変動は話者によって異なっているものの、フォルマンの位置関係を観察すると、3 kHz 以上のフォルマンがより影響を受けていることが分かる。特に話者 2 ではその影響が顕著であったことが見てとれる。図 8 の鼻音 /n/ においても同じ傾向が見られた。

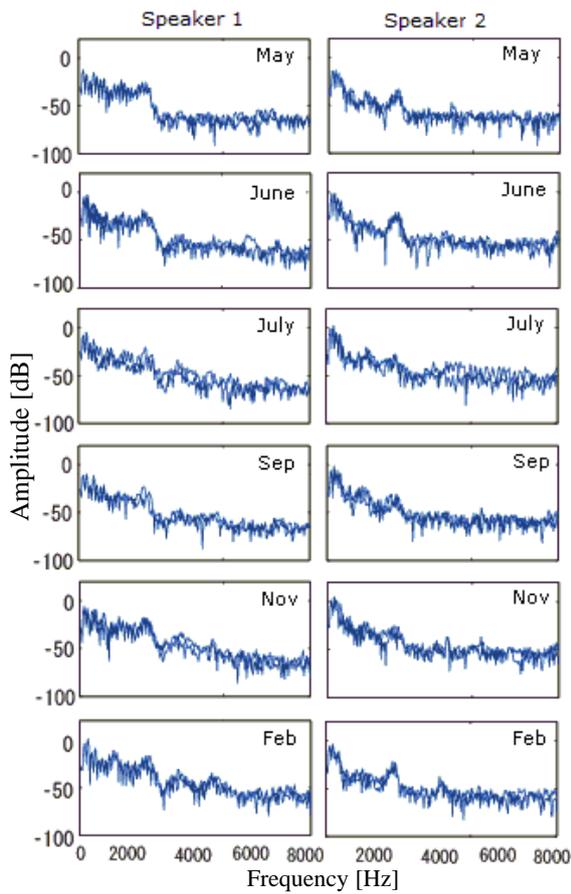


図 3. 話者 2 名による連続発話中の /m/ (m1)

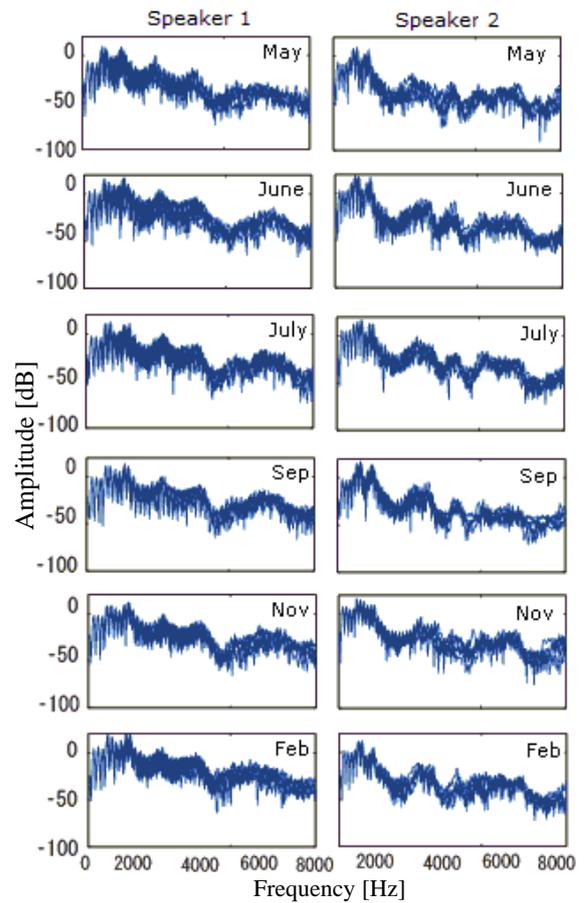


図 5. 話者 2 名による単独発話母音 /a/

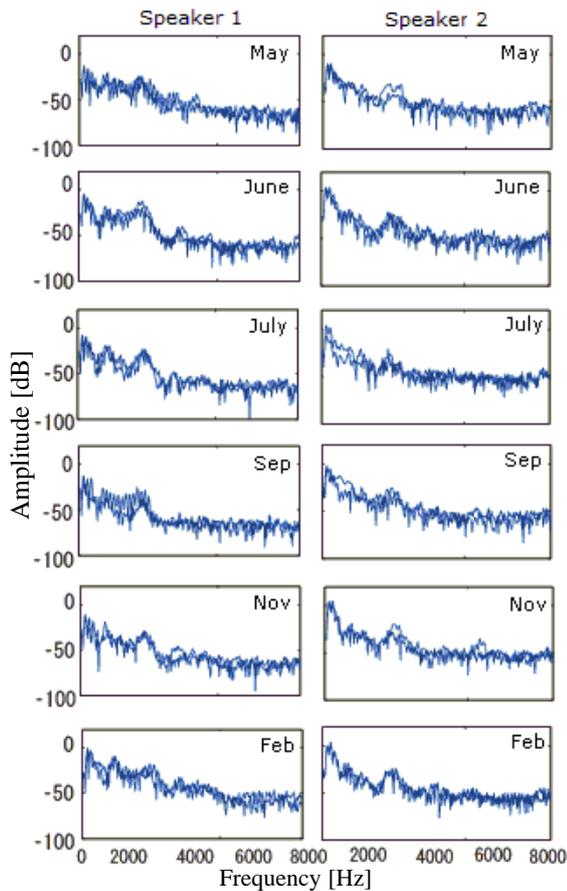


図 4. 話者 2 名による連続発話中の /m/ (m2)

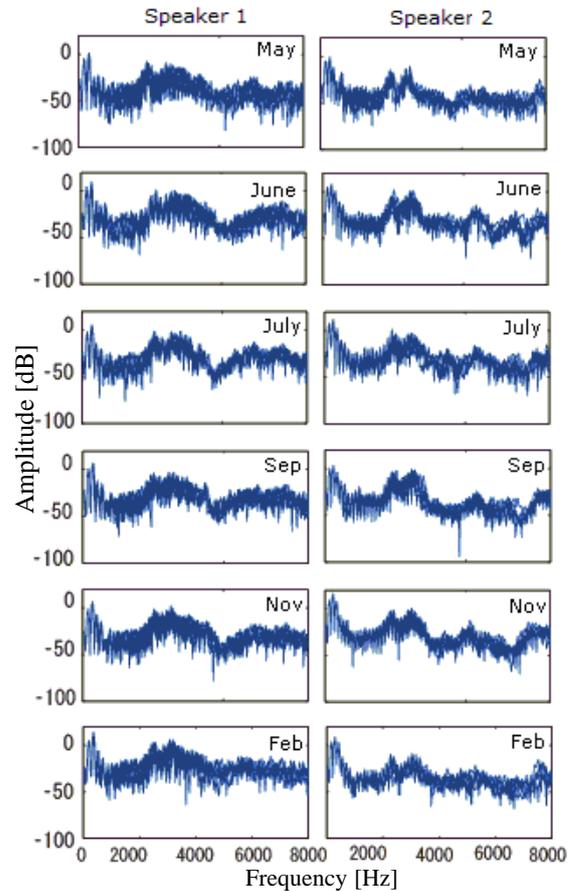


図 6. 話者 2 名による単独発話母音 /i/

#### 4 まとめと考察

話者 2 名から 10 ヶ月に渡って録音した連続発話および単独発話の音声进行分析し、スペクトルの話者性が経時的にどのように変化するかを調査した。連続発話から抜粋した音声では、経時変化よりも前後の文脈（音環境）の影響が大きいことが分かった。単独発話音声では、音素によって話者間の違いの現れ方が異なっており、経時変化の現れ方も話者や音素によって異なることが分かった。

今後はより多くの話者の音声进行分析する必要がある。特に鼻音は鼻風邪や花粉症によって大きく影響されるため、話者の健康状態も考慮に入れる必要がある [27-28]。

また音源情報は 10 年、20 年と長期的に見ると大きく変化することが分かっており [28-29]、今後は声道情報に焦点をあてて、スペクトル包絡を観察するのがよいと思われる。

#### 謝辞

この研究は文部科学省私立大学学術研究高度化推進事業 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] Amino *et al.*, *Acoust. Sci. Tech.*, 27(4), 233-235, 2006.
- [2] Amino&Arai, *Acoust. Sci. Tech.*, 28(2), 128-130, 2007.
- [3] Amino *et al.*, *IEICE Tech. Rep.*, SP105(685), 109-114, 2006.
- [4] Amino *et al.*, in “Speaker Classification 2,” 83-92, Springer, 2007.
- [5] Amino&Arai, *Acoust. Sci. Tech.*, 30(2), 89-99, 2009.
- [6] Amino&Arai, *IEICE Tech. Rep.*, SP107(551), 29-34, 2008.
- [7] Amino&Arai, *Proc. Acoustics '08*, 2381-2386, 2008.
- [8] 中川&坂井, *音響誌*, 35(3), 111-117, 1979.
- [9] O’Shaughnessy, “Speech Communication,” Addison-Wesley, 2000.
- [10] Fattah *et al.*, *IEICE Trans. Info. Sys.*, E89-D, 1712-1719, 2006.
- [11] 波々伯部ら, *音講論(秋)*, 715-716, 2008.
- [12] 奥ら, *音講論(秋)*, 119-120, 2009.
- [13] Park&Hazen, *Proc. ICSLP*, 1337-1340, 2002.
- [14] Sigmund, *J. Elect. Eng.*, 57, 301-303, 2006.
- [15] Lei&Lopez, *Proc. Interspeech*, 2323-2326, 2009.
- [16] 古井, “デジタル音声処理,” 東海大学出版会, 1985.
- [17] 古井, *信学論*, J57-A(12), 880-887, 1974.
- [18] 古井, *信学論*, J65-A(2), 183-190, 1982.
- [19] 高梨&柏木, *信学論*, J74-A(12), 1810-1811, 1991.
- [20] 松井&古井, *信学論*, J75-A(4), 703-709, 1992.
- [21] 松井&古井, *信学論*, J77-A(4), 601-606, 1994.
- [22] 早川ら, *信学技報*, SP, 95(319), 17-24, 1995.
- [23] 松井ら, *信学論*, J81-D2(2), 268-276, 1998.
- [24] 柘植ら, *音講論(春)*, 129-130, 2005.
- [25] 河井&津崎, *信学技報*, SP102(292), 1-4, 2002.
- [26] 山田ら, *音講論(秋)*, 195-196, 2009.
- [27] 網野&荒井, *音講論(秋)*, 431-432, 2005.
- [28] 網野&荒井, *音講論(春)*, 363-364, 2006.
- [29] 鈴木ら, *信学会大会(情報システム)論文集*, 337, 1998.
- [30] 西尾&新美, *音声言語医学*, 46, 136-144, 2005.

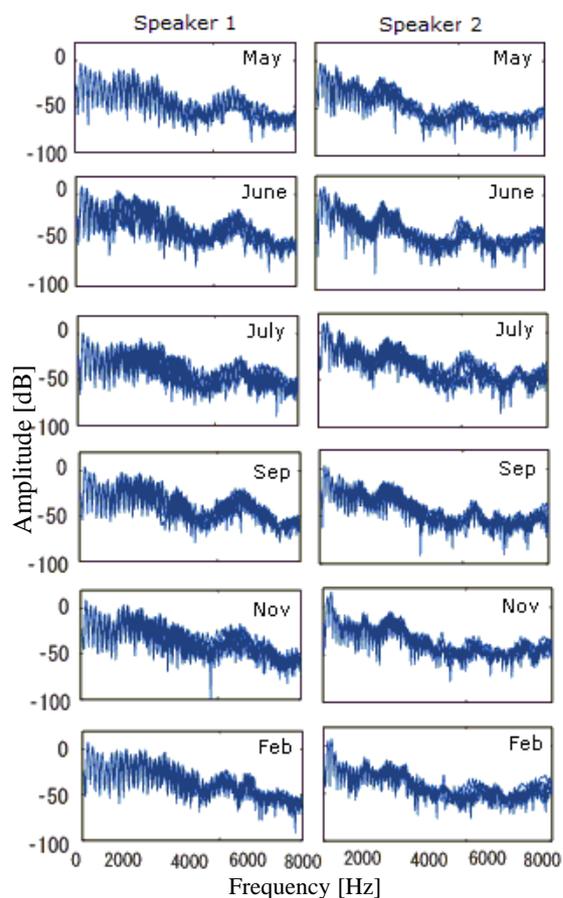


図 7. 話者 2 名による単独発話の鼻音 /m/

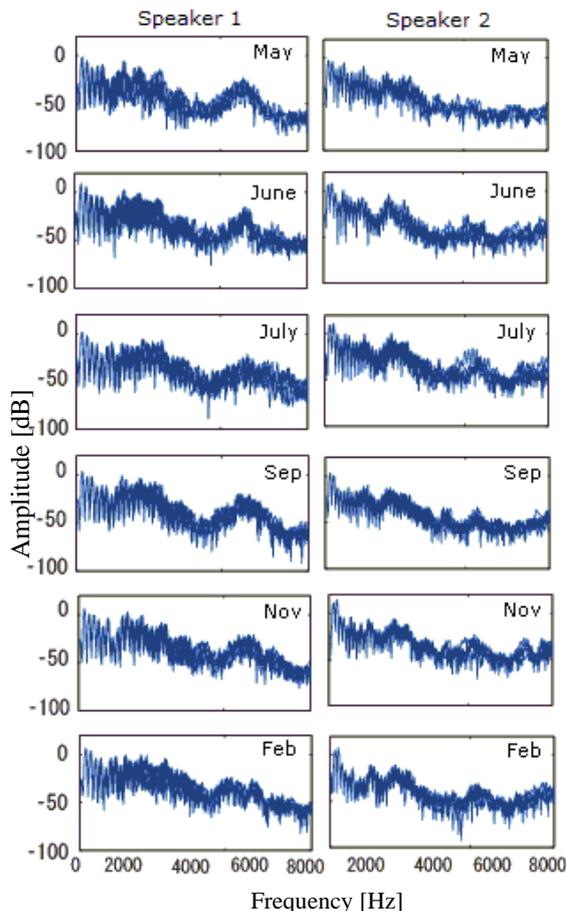


図 8. 話者 2 名による単独発話の鼻音 /n/