HOKKAIDO UNIVERSITY

Title	九官鳥母音の音声認識について
Author(s)	新保, 勝; 伊達, 惇; 宮腰, 政明
Citation	北海道大學工學部研究報告, 99, 103-108
Issue Date	1980-08-11
Doc URL	https://hdl.handle.net/2115/41621
Type	departmental bulletin paper
File Information	99_103-108.pdf

Hokkaido University Collection of Scholarly and Academic Papers : HUSCAP

北海道大学工学部研究報告
第99号 (昭和55年)

Bulletin of the Faculty of Engineering,
Hokkaido University, No. 99 (1980)

九官鳥母音の音声認識について

新保 勝* 伊達 惇** 宮腰政明*

(昭和55年3月31日受理)

On the Phonetical Recognition of Mynah Mimic Vowels

Masaru SHIMBO Tsutomu DATE Masaaki MIYAKOSHI

(Received March 31, 1980)

Abstract

It is justified to classify the mynah mimic vowels as well as human vowels from the standpoint of the mathematical phonetical theory proposed elsewhere, where three, and only three, effective formants are accorded to the vowel structure. In the present paper mynah's and her tutor's vowels are analyzed by a soundspectrograph to extract formant frequencies, and vowel phonemes are plotted on a projective vowel plane with axes of the logarithm of formant frequency ratios. Though the present data are restricted and few, the separation of phonemes are rather good and there can not be found a remarkable difference between mynah's and her tutor's vowel clouds.

1. まえがき

九官鳥の物真似はよく知られている。九官鳥の発生器官は人間と異なることから，発声機序についてはこれ迄に声道共振特性説[1),2)]や鼓状膜の振動説[3)]があり，更には音声認知機序については第1フォルマントのレベルが小さいことから，第2フォルマントが音韻識別の鍵である[4)]とする考えや，或は，人間の聴覚内の非線形処理に基づくとの示唆[1),2),5)]がなされているが，いずれも実証されておらず，まだ仮説の段階にある。

他方で人間の3フォルマント母音については，周波数の対数差を座標とする射影的母音平面上で識別できることは既に指摘され，実験的にもこの分類関係がよくなることが確かめられている[6),7),8),9),10),11),12)]。本稿ではこの理論を九官鳥の母音に適用し，人間の母音感覚における認知機序解明の手懸りを求める。

2. 九官鳥音声の分析

本報告は，九官鳥について特別な手法を用いたということではなく，通常の人間の母音分析の方法[13),14),15),16)]によって解析を行った結果に関する報告である。対象とする九官鳥の音声は，日本音響学会音声研究会資料[5)] (S 79-66) および聴覚研究会資料[5)] (H-68-2) において使用された音声テープを借用したものであり，さらに同資料で使用したパワースペクトル図の一部も借用した。

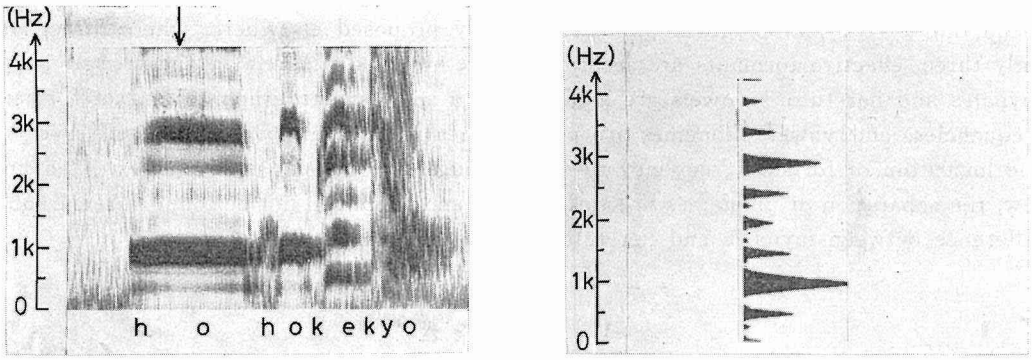
* 情報工学専攻 情報処理工学講座
** 情報科学教室

手順は，前者の音声テープからサウンドスペクトログラムを作成してフォルマント周波数を読みとること，後者からは直接フォルマントを読みとること，の両方である。基礎データである九官鳥およびヒトの音声は12単語であり，耳で聞いた場合，かなり明瞭で音量もあり，分析対象として特別の工夫を要するというものではなかった。したがって，原テープを通常のテープレコーダーで再生し，サウンドスペクトログラフ（リオン社製 SG 07 型）に接続し，サウンドスペクトログラムパターンおよびセクションパターンを作成し，その両方を用いて母音に関するフォルマント周波数を読みとるという方法に依った。

基礎データである九官鳥の発声のうち，分析の対

表—1　分析の対象とした音声

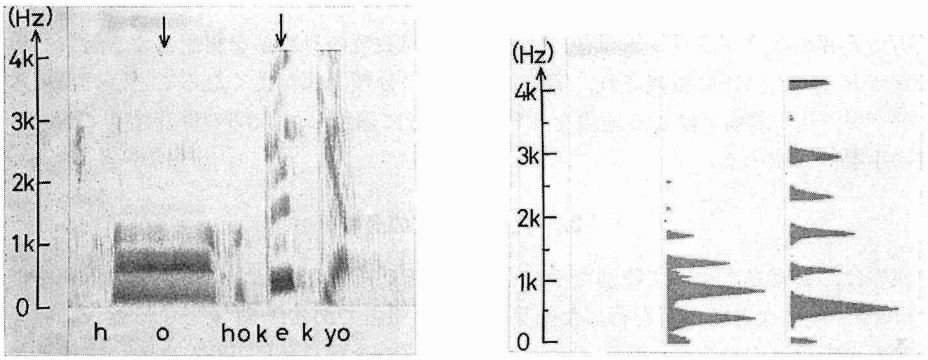
母　音	単　語
/a/	bai bai onichan gomenkudasai
/i/	konnichiwa onichan kawaiine
/u/	kyuchan
/e/	kawaiine gomenkudasai
/o/	popopo hohokekyo



(a)　　(b)

図—1　九官鳥のサウンドスペクトログラム (hohokekyo)

(a)　広帯域パターン
(b)　狭帯域セクションパターン



(a)　　(b)

図—2　ヒトのサウンドスペクトログラム (hohokekyo)

(a)　広帯域パターン
(b)　狭帯域セクションパターン

象となった単語を表-1に記す。また，サウンドスペクトログラムパターンおよびセクションパターンの一例の写真を図-1および図-2に示す。図-1(a)に示されたサウンドスペクトログラムパターンは，縦軸が周波数 (0～4,000 Hz)，横軸が時間 (0～2.4 秒) を表わし，図中の黒さの濃度が対応する周波数成分の強さを表わしている。使用している濾波器の帯域幅は 300 Hz である。次に，図-1(b)のセクションパターンは，図-1(a)のパターンの時間軸上の点を指定し，その時刻における，各周波数成分の強さを横軸として表したものである。その際分析に使用された濾波器の帯域幅は 45 Hz である。

作成したサウンドスペクトログラムに共通して言えることとして，九官鳥では高周波成分が強調されるという通説が首肯されること，しかし，第1フォルマントは存在しないという程ではなく，相対的に弱いけれどもその存在は否定しえないことが挙げられる[3),4)]。また，九官鳥のピッチ周波数はかなり高く，300～500 Hz ぐらいであった。これは，本来的なものか Tutor となったヒトの声が女声であったことによるものかは，これだけのデータからその理由を明らかにすることはできない。一般にピッチ周波数が高い音声の母音フォルマント分析は容易ではなく，本件もその例外ではない。

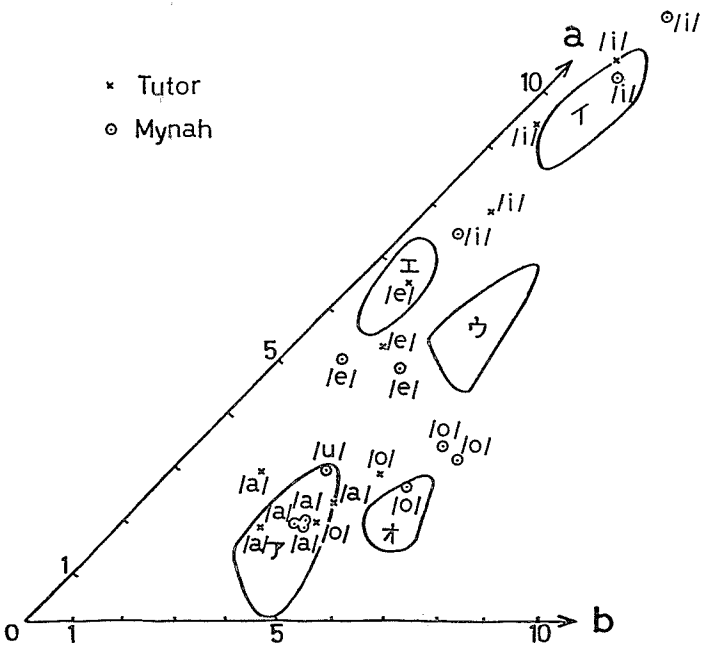
表-1に記した単語から作成したサウンドスペクトログラムパターンおよびセクションパターンによって読みとったフォルマント周波数を表-2に記す。得られたフォルマント周波数から，第3節で詳述するパラメータ a, b を計算して，各音韻を (a, b) 平面上にプロットしたものが図-3である。

図-3における a, b 両軸の角度は45度になっているが，この角度は別に意味はなく，便宜的なものである。図において，各母音を示す領域を囲った閉曲線は文献6)から借用したものであり，×印のわきにアルファベットで母音を示してあるのが今回の測定計算の結果得られたヒトの音声を表わし，◉印のわきにアルファベットで母音を示してあるのが九官鳥の音声である。

個別に観察を行うと，/a/ については，九官鳥もヒトも割合にまとまっており，文献6)で得ら

表—2 フォルマント周波数

	Mynah			Tutor		
	F_1 (Hz)	F_2 (Hz)	F_3 (Hz)	F_1 (Hz)	F_2 (Hz)	F_3 (Hz)
/a/	650	1,080	2,630	620	1,050	2,550
	750	1,150	2,600	650	1,000	1,920
	750	1,180	2,700	850	1,660	2,520
/i/	500	2,740	3,500	350	3,060	3,480
	250	2,650	3,500	450	2,750	3,720
	200	2,800	3,470	280	3,220	3,930
/u/	500	980	1,980			
/e/	800	2,550	3,350	530	2,350	3,000
	650	1,970	3,500	620	2,080	3,130
/o/	590	1,070	3,360	600	950	2,350
	450	920	2,850	500	960	2,520
	450	970	2,980			



図—3 音声母音三角形における九官鳥およびヒトの音韻分布
($a=10\log_{10}F_2/F_1$, $b=10\log_{10}F_3/F_2$)

れた標準の音韻領域によく一致する。とくに九官鳥の場合は，点の位置のバラツキも少ない。/i/ および /e/ については，九官鳥もヒトもデータ間のバラツキが大きくなっており，その位置も標準母音領域からずれたところにはみ出ている。しかし，標準母音領域が多少広がっていると考えれば説明が可能，という程度である。次に，/o/ については，九官鳥もヒトも標準母音領域の周辺に散在するという結果になっているが，ヒトの方に /a/ の標準母音領域へ入り込むデータがあって，九官鳥の方がむしろ識別が容易という現象が起きている。/u/ については，データが /kyuchan/ の /u/ がたった一つあるのみであり，ヒトの場合，第2，第3フォルマントの読みとりが困難であったため，図-3にはヒトの /u/ が記入できなかった。九官鳥の場合も，他のデータに比して読みとりにくかったが，ヒト程ではないので作図し，その結果，/a/ の標準母音領域に入り込んだものが得られた。

以上が，1羽の九官鳥の発声する12単語の母音部分のサウンドスペクトログラム分析の直接的な結果である。図-3の解釈は第3節にて行うが，データそれ自体が少ないこと，フォルマント読みとりが必ずしも一意的かつ不動なものではないこと，などの理由によって，限界があることは指摘されよう。

次に，こうした限界を補完する目的で，前述の音声研究会資料[5]において使用されたパワースペクトラム（約50 μsecでサンプルした1,024点に対し，46 m secのハニング窓をかけ，F. F. T. を用いて算出したとされる）を借用し，それらから直接的にフォルマントを読みとろうとしたが，その結果は，母音平面の /a/, /o/, /u/, /e/ で囲った領域に各音韻が集中することとなり，採用を断念した。中間データそれ自体は貴重であるので，その性質を十分解明した上で，再度活用を試みる所存である。

3. 母音音声の心理物理的認識

音声感覚の数理的解析に基いて，フォルマント周波数の対数差を座標とする射影的母音平面上で母音，変母音等を識別できる可能性は既に報告され[6),7),8),9),10),11),12)]，実験的にも確められているが，そこで仮定された母音認識の基本構造は

1) 3フォルマント構造である，

2) フォルマント周波数の対数差が有効である，

3) フォルマントの強弱は寄与しない，

に基礎をおくものである[7)]。

仮定1)は母音感覚が3個のフォルマントで決定され，残りのフォルマントは副次的であるとするものである。これはヒトに限らず，九官鳥の場合にも適用され得るものであろう。更には仮定3)を考慮すれば，九官鳥に特有なフォルマントの高次成分の強調[2)]は音韻分類に本質的な意味を持たないことになる。同じく九官鳥で第1フォルマントに相当するものが弱いけれども，存在し[3)]，母音として認識できることはフォルマントの強弱によらぬとする仮定3)の実証とも考えられる。また，第2フォルマントが音韻識別の鍵[4)]であるとするのも，仮定1), 3)の立場から想定されるものとは異なる。

仮定2)は心理物理感覚の基本原理であるウェーバー・フェヒナー則に関係して，母音分化に有効なのはフォルマント周波数そのものでなく，その対数差であるとの主張である。これからフォルマント周波数比の対数が本質的に意味のある量となり，聴覚は二つの周波数の距離の対数を識別することになる。あるいは，射影幾何学的見地からすると，非調和比（複比）の対数であり，周波数が3個ということから，母音感覚の集合は二次元射影平面をなすといえる。したがって母音のパラメータとして，例えば

$$a = 10\log_{10}(F_2/F_1), \qquad b = 10\log_{10}(F_3/F_2) \qquad (F_1 \leqq F_2 \leqq F_3)$$

を取ることにより，人間ならびに九官鳥の母音をパラメータ a, b を座標軸とする母音平面上にプロットすることの理論的根拠を与えることになる。ここで F_1, F_2, F_3 は母音の第1，第2，第3フォルマント周波数である。なお，座標軸の方向は母音感覚の射影性から適当に選んであり，大略，a 軸は鋭鈍，b 軸は軽重の度合[17)]を示している。

前節で分析した九官鳥の母音は1羽による12単語であり，少数の限られたデータであるとともに，フォルマント周波数の抽出に不確実さが残ることを注意しなければならない。こうした場合，母音はズレあるいはバラツキを伴って標準母音領域外に生ずる可能性が十分考えられる。しかし，/u/を除き，全体的にヒトと九官鳥の母音間にそれ程著しい差が見られないことを考慮すると，母音の3フォルマント構造の仮定の妥当性やフォルマント周波数が弱いにもかかわらず，読取は可能であり，図-3の母音平面上で概ね分離可能であることはこの分類法に対して，実験的にも前向きな支持を与えるものと思われる。言いかえれば，ヒトと九官鳥における発声器官の構造と機能は音声の種類に制限を加えるが，それ以上にこの認識方法は不変な意味を持つように見受けられる。

4. あ と が き

音響学会で報告されたヒトと九官鳥の音声分析データー[1),2)]からサウンドスペクトログラムを作成し，これから読みとった3個のフォルマント周波数の対数差をパラメータとして，発声母音を射影的母音平面上で分類することを試みた。用いたデータは少数の限られたものであるが，

音韻分離は概ね可能であり，ヒトと九官鳥の間に顕著な差は見出せなかった。このことから，発声器官の構造や機能の差以上にこの分類法は不変な意味を持つように見受けられる。九官鳥フォルマントの発生機序やヒトの聴覚内での非線形処理[1),2),5)]に関しては特にふれなかったが，フォルマント抽出法とともに今後更に検証される必要があろう。

この研究にあたり，本学応用電気研究所吉本千禎教授，平原達也氏にはヒトと九官鳥の発声母音に関する貴重な実験資料をお貸し頂いた。また，母音スペクトルの分析にあたり，本学医学部本間行彦講師の御援助を頂いた。ここに厚く謝意を表する次第である。

文　　献

1) 平原達也，伊福部達，吉本千禎： 日本音響学会講演論文集 (1978), 315.
2) 平原達也，伊福部達，吉本千禎： 医用電子と生体工学, **17** (1979).
3) D. H. Klatt and R. A. Stefanski: J. A. S. A., **55** (1974), 822.
4) 稲富和夫： 日本音響学会講演論文集 (昭 46-11), 385.
5) 平原達也，伊福部達，吉本千禎： 日本音響学会音声研究会，S 79-66／聴覚研究会資料，H-68-2 (1980), 505.
6) M. Iri: RAAG Memoirs, **3** (1962), 551.
7) 近藤一夫： 数理音声学序説，東京大学出版会 (1964), 8.
8) K. Kondo *et al.*: RAAG Memoirs, Division H; **3** (1962), 511; **4** (1968), 463.
9) K. Kondo and M. Shimbo: RAAG Memoirs, **4**, H-VIII (1968), 467.
10) T. Date: RAAG Memoirs, **4**, H-XII (1968), 522.
11) M. Shimbo: Bull. Fac. Eng., Hokkaido Univ., **77** (1975), 145.
12) M. Shimbo and K. Kondo: Prevenient Natural Philosophy, **93** (1979), 9.
13) S. S. Stevens, J. P. Egan and G. A. Miller: J. A. S. A., **19**, 5 (1947), 771.
14) J. Lotz: J. A. S. A., **22**, 6 (1950), 712.
15) J. L. Flanagan: Speech Analysis, Synthesis and Perception, Springer-Verlag (1965).
16) G. Fant: Speech Sounds and Features, The MIT Press (1973), 143.