

論文 / 著書情報
Article / Book Information

論題(和文)	話者の個人性に対する聴覚の適応特性
Title	
著者(和文)	加藤 和美, 古井 貞熙
Author	SADAOKI FURUI
出典(和文)	騒音研究会資料(H85-01-5), , , pp. 1-8
Journal/Book name	, , , pp. 1-8
発行日 / Issue date	1985, 1

話者の個人性に対する聴覚の適応特性

Listener Adaptability for Individual Voice in Speech Perception

加藤 和美 古井 貞熙

Kazumi KATO Sadaoki FURUI

日本電信電話公社 武蔵野電気通信研究所

Musashino Electrical Communication Laboratory, N.T.T.

人間が音声を認識する場合の能力について明瞭度を用いて実験的検討をおこなった。多数話者の音声の混合呈示による認識実験の結果、話者に対する適応は4-5音節程度の音節の受聴で達成されることがわかった。さらに適応特性を音素別に分析し、破裂音に対しては適応効果が大きく、拗音では適応効果の小さいことを明らかにした。また、調音様式、調音部位の違いによる効果の差を求めた。話者の違いによる影響については、適応前の明瞭度の低い話者の方が適応効果の大きいことがわかった。つぎに適応された状態が異話者の音声の受聴によって適応前の状態に戻る適応の消滅特性について実験を行った結果、6-7音節の異話者の音節の受聴によって適応が失われることがわかった。

1. はじめに

音声にはその意味内容に関する情報のほか、話者の個人性情報や喜怒哀楽に関する情報など、種々の情報が含まれている。このうち意味内容に関する情報が第一義的であることは言うまでもないが、その情報は通常さまざまな要因によって変形している。その要因には、周囲の雑音、残響の付加などの外的ひずみと、音声レベルの大小、話者の個人性、その声質の時期的あるいは発声のたびごとの変化などの内的ひずみがある。人間は何の気なしに音声を聴いていても、これらのさまざまな音声の変形に対して高度の適応能力を発揮しており、特に声質の個人差に対する適応能力は重要な機能の一つである。声質のパラエティーには、子供の声、年寄りの声、張りのある声、かすれ声、明るい声、暗い声、笑い声、など数えあげればきりが無い。これらの種々の声に対して誰でも、初めての人や思いがけない人に話しかけられると最初何を言われたかはっきりわからなくても、普通の場合はすぐにその人の声に慣れて（適応して）よくわかるようになる。このとき、人間は話者の声質の特徴を、音声のどのような特徴を手掛かりとして、どのように獲得するのであろうか。また、聴覚においてはこのときどのような情報処理がなされているのであろうか。ここでは、話者の個人性への適応特性に着目し、多数話者の音声を聴取

する場合の聴覚機構について検討を行った。

現在、音声の機械認識は広く研究され、一部の領域では実用にも供されている。しかしこれらが認識できる音声は、あらかじめ声を登録した話者が発声したもので、音節単位あるいは単語単位で発声したものである必要があり、不特定多数の音声を対象とするものは認識できる語数が極めて限られている。大語彙、不特定話者に対する認識技術に関しても多くの研究がなされているが、まだ未解決の問題が多い。現行の機械認識における、さまざまな音声の変形要因に対する適応能力は、人間の能力には遠く及ばない。人間のそれらに対する音韻認識能力を明らかにすることは、聴覚における情報処理機構の解明に役立つと同時に、機械認識技術の進歩への示唆を与えられよう。

2. 話者の個人性に対する適応のモデルと実験の目的

声質の個人性に対する適応のメカニズムとして、一例として図1に示すようなモデルを考えることができよう。すなわち、音声刺激として聴覚器官に捕らえられると、音韻認識が行われると同時に声の質、即ち話者の音韻的特徴の抽出が行われる。このとき、音韻の認識と話者特徴の抽出は相互に密接に関連しており、音韻がわかることによって話者特徴が抽出でき、話者特徴がわかることによって音韻が認識しやすくな

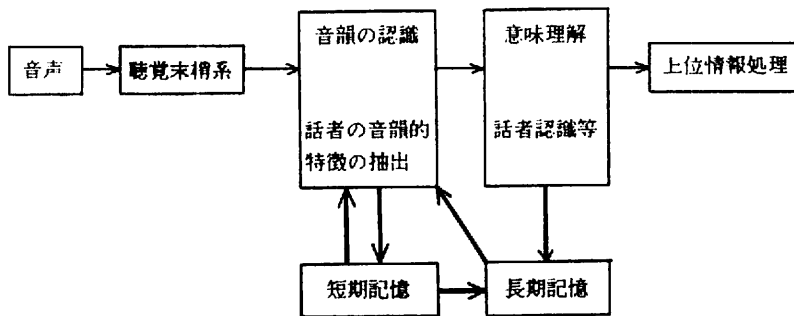


図1 聴覚情報処理機構

るという関係にある。聴取者の長期記憶の中にその話者についての声の特徴がすでに記憶されていて利用できる場合は、ただちにそれが音韻の認識に役立つであろう。ところが、初めて聴く人の声であるとなると、最初のステップでは、その聴取者の持っている一般的な音声に対する情報が利用されて音韻認識が行われ、同時にその話者の音声の音韻的特徴に関する情報が短期記憶に記憶される。そしてその特徴が次の音声の認識に利用されていく。このような処理が連続して行われると、音韻の認識はその話者の特徴が十分に生かされたレベルで処理されるようになり、聴覚的にその話者の音声に適応した状態となる。このような能力は一般に経験されていることあり、種々の話者の音声を混ぜて呈示したときと、同一話者の音声だけをまとめて呈示したときとで、母音の正聴率が異なることはいくつかの実験ですでに確認されている[1,2,3]。しかし、その効果が子音を含む音節に対してどの程度であるか、どのように行われるかなどについてはまだ明らかになっていない。

ここでは高度な意味・言語情報を持たない音韻レベルでの適応能力について次に示す明瞭度を用いて検討した。即ち、音声を構成する基本単位である単音節100個を用い、種々の話者条件において、受聴した音節のうち正しく認識できた数、いわゆる単音明瞭度、子音明瞭度、母音明瞭度、子音正聴率を用いて評価を行った。明瞭度試験を行うにあたっては試験法自体の持つ問題点について注意を払い、他の要因の影響をなるべく排除するよう努めた[4]。以下特に断らないかぎり、明瞭度とは単音明瞭度をさすものとする。

3. 音韻認識実験の方法

表1および図2に示す実験条件のもとで認識実験を実施した。詳細は前回報告を参照されたい。

4. 実験結果と考察

表1 実験条件

音声資料	話者 男性:46名 女性:54名 音声 日本語100 音節、帯域:8kHz パワー 音声電圧計の振れがほぼ一定
受聴者	通話試験員 女性4名
系条件	日本主通話標準装置 線路損失:52dB Hoth 雑音:1mV

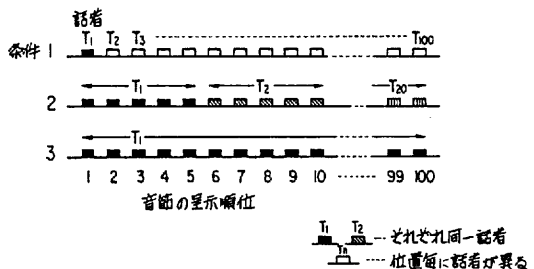


図2 音節の呈示条件

4.1. 話者の特徴が利用されない場合の音韻認識能力(呈示条件1)

呈示条件1のように話者が1音節毎に異なると、その話者に関する情報は次の音韻の認識に全く利用できない。この実験の結果から次のことが明らかとなった。

(1) 実験期間(約1ヶ月)を通じて100音節リスト毎の明瞭度値が上昇していく傾向は見られなかった。このことは、直前の音節から発声者に対する知識が獲得できない状況において、100名の個々の話者の音声

の特徴を徐々に長期記憶にとどめて、それを検索することによって音韻認識能力を向上させるというような働きは存在しないことをしめている。この場合、音声の話者特徴についての情報が短期記憶だけに留まったか、あるいは長期記憶に記憶されたが有効に利用されなかったについては明らかではない。

(2) 話者別の明瞭度の、100名の平均は71.6% (男声は70.5%, 女声は72.6%) で、ばらつきは、 $\sigma = 6.1\%$ であった。

(3) 100名の話者について年齢別の明瞭度を図3に示す。統計的に有意ではないが男女とも若年層の明瞭度は比較的低く、30才以上となるとほぼ一定となる傾向がある。

4.2. 話者の特徴を利用できる場合の音韻認識能力 (呈示条件3)

連続100音節を同一話者とした場合 (呈示条件3) の結果を前項の呈示条件1の場合と合わせて表2に示す。一般人、試験員のいずれの場合でも、呈示条件3の方が呈示条件1よりも明瞭度が高くなっている。このことは、同一発声者の音節を続けて聴くことによって、発声者の特徴を記憶して次の音韻の認識に利用していく機構が存在することを示している。

4.3. 適応能力の分析

4.3.1. 音節の呈示数の効果

呈示条件2において、同一話者が発声する5音節の呈示位置毎 (1~5) に全話者を平均した明瞭度を求め、呈示条件1および3の結果と合わせて図4に示した。図4からわかるように、呈示条件2の位置1では、ほぼ呈示条件1と同程度の明瞭度を示し、位置2から5へ同一話者の音節の呈示が進むにつれて、明瞭度が上昇していく傾向がある。またその値は4~5音節程度で飽和していることがわかる。このことから人間は、4~5音節程度を連続して聴けばその話者についての音声の音韻的特徴を十分捕らえ利用することができるといえる。

斎藤[7]は明瞭度試験において、100音節の送話者数が1および4人の場合について、音韻識別に及ぼす送話者数の影響を検討している。その実験結果では、今回の実験と同じ帯域制限のない場合については、送話者が1名であっても4名であっても、その人数の影響はないとしている。これは (1) 送話者は通話試験員であり日常その100音節を聴き慣れていること、(2) 送話者が4名の場合でも1人が25音節を連続して発声している。これらのことから話者に対して十分適応してしまい、1人が100音節を発声した場合と差が生

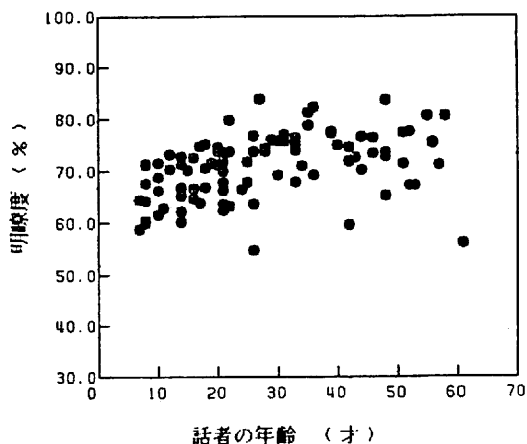


図3 話者別の明瞭度 (呈示条件1)

表2 呈示条件と音韻認識結果 (明瞭度)

条件		呈示条件1		呈示条件3	
		子音	単音	子音	単音
話者	一般人 (30名)	46.20	70.51	51.95	75.21
	試験員 (5名)	56.80	75.75	65.05	81.17

(単位: %)

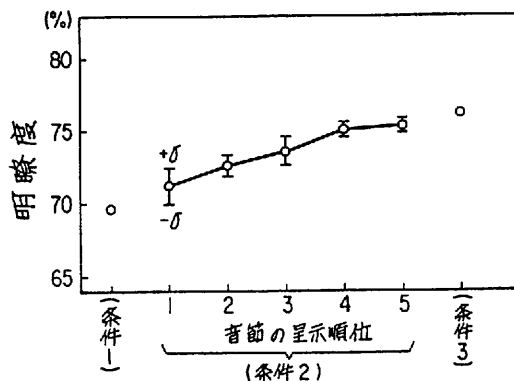


図4 呈示位置に対する適応特性

じなかったのではないかと考えられる。

4.3.2. 呈示音素 (子音) の効果

日本語の音節は一般に (子音C+母音V) で構成さ

れている。話者への適応に利用していると考えられる5音節の中に同一の音節が2つ以上出現することはないが、母音はもちろん、同じ子音の出現する可能性はある。子音のうち100音節中に1回しか出現しないものは/w/. /ts/であり、3回出現する音素は、/t/. /d/. /j/. /pj/. /bj/. /kj/. /gj/. /nj/. /mj/. /hj/. /rj/. 4回出現するものは、/s/. /ʃ/. /dʒ/. /dz/. /tʃ/であり、/p/. /b/. /k/. /g/. /m/. /n/. /h/. /r/は5回出現する。

呈示条件2の呈示順序を分析し、同じ音素(子音)が適応過程の該呈示位置よりも前に1回以上出現した場合と1回も出現しなかった場合で、適応効果に影響が現われるかどうかを検討した。

図5に分析結果を示す。図中横軸は呈示位置、縦軸は明瞭度である。位置毎のデータが少ないためばらつきが大きい。これによると適応過程に同種の音素が呈示されてもされなくても、結果には有意差がない。音節の異聴は主として子音に関して起っており、同じ子音が適応過程に呈示されなくても適応が進んで、子音の異聴が減少するということは、母音に依存して子音の適応がなされていることを示唆していると言える。人間による子音の受聴は、主として母音とのまどまり即ち音節として行われ、基本的に母音部への遷移として子音が受聴されることが、別の実験で確認されている[8]。このために母音の個人差への適応が間接的に子音の正聴率を高めることになると思われる。

4.3.3. 音素別の適応効果

十分に話者に対して適応した場合(呈示条件1)と、適応前(呈示条件3)の各子音の正聴率の差、即ち適応の大きさを図6に示す。図6には、さらに後続母音の種別による効果の差も示してある。これからわかるように、適応効果は音素によってかなりの差があり、/t/. /k/. /d/. /g/などの破裂音で効果が高い、すなわち正聴率の向上が著しく、/w/. /j/などの半母音、/r/などの弾音、拗音などでは効果が低い。

そこで調音様式と調音部位別に効果を求めてみたところ、図7が得られた。調音様式別では半母音、弾音で極端に効果が小さい。また調音部位別では、硬口蓋音が効果が見られないことが示されている。図6からわかるように、後続母音別ではほとんど差はないが、/i/の場合にやや効果が大きく、/a/の場合に効果が小さい。この結果は、母音の個人差は前舌母音で比較的大きく後舌母音で比較的小さいという過去の知見と対応している[9]。なお、適応前の各音素の絶対値と正聴率の向上効果との間には有意な相関は見られなかった。

4.3.4. 適応による母音の異聴の変化

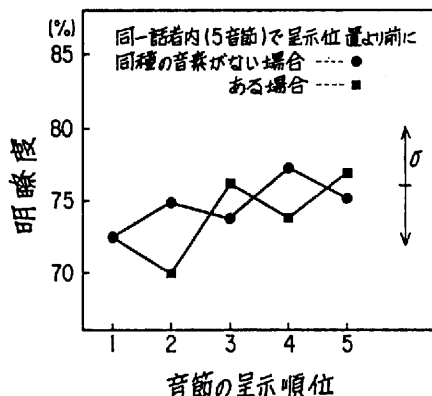


図5 同種の音素呈示の影響

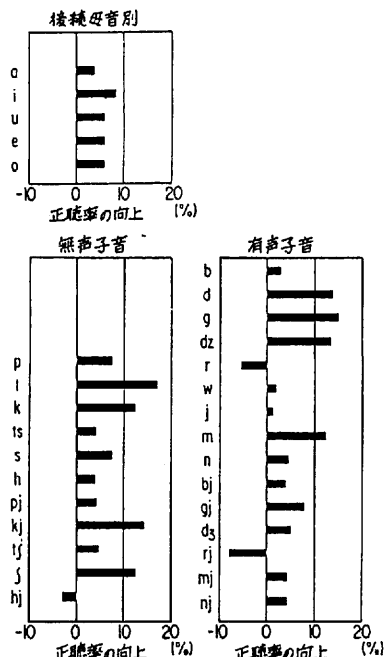


図6 適応による各音素の正聴率の向上

表3に適応前と適応後の母音異聴表を示す。全体としての母音明瞭度は、94.8%から98.5%へと3.7%向上し、母音別には適応前の明瞭度が一番低い/u/が90.5%から96.3%へと5.8%向上する。/a/は適応前にすでに99.1%に達しているため、0.8%しか向上しない。

異聴では、/u/→/i/が4.5%とかなり大きく、適応しても2.6%の異聴が残る。また/i/→/u/の異聴は、適応前には5.0%であるが適応すると1.2%に改善される。

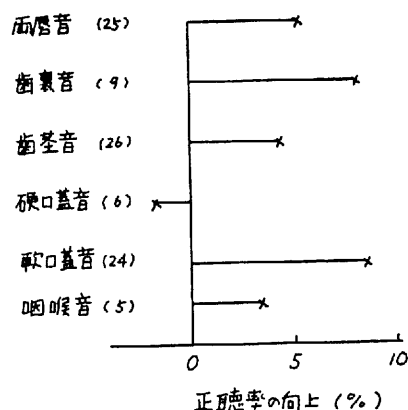
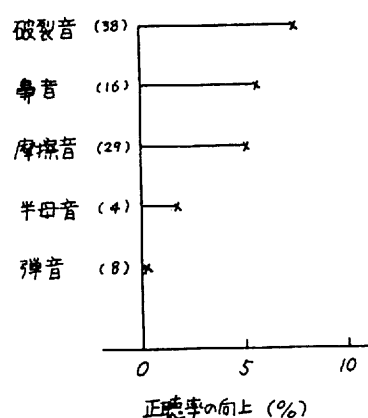


図7 調音様式、調音部位別の適応効果

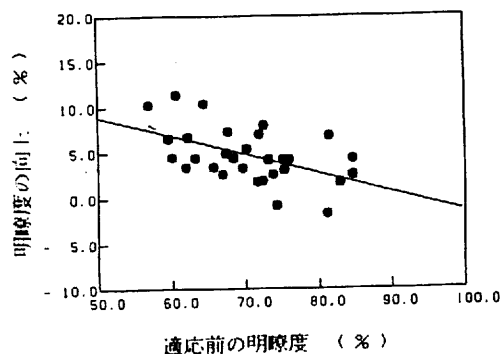


図8 話者による適応効果の違い

表3 母音異読表 (単位 %)

	適応前					適応後				
	i	e	a	o	u	i	e	a	o	u
i	94.5	0.4	-	0.1	5.0	98.8	-	-	-	1.2
e	0.3	93.9	-	3.4	2.4	0.5	97.8	-	1.0	0.6
a	-	0.1	99.1	0.7	0.1	-	-	99.9	0.1	-
o	0.2	1.3	2.5	95.4	0.6	-	0.4	0.2	99.3	0.1
u	4.5	1.5	0.6	2.9	90.5	2.6	0.2	0.1	0.7	96.3

4.3.5. 話者の差と音韻認識能力

いままでの結果は話者間の差は考えずに平均した値で示してきたが、実際には、話者によって適応効果にかなりの差がある。図8は送話者30名について、適応前の明瞭度と適応効果を示したもので、両者の相関係数は -0.525 である。適応効果は、 -1.6% から $+11.5\%$ まで変動している。ここで話者を3つの年齢層にわけ、年齢層毎に対応をとって見たところ、図9が得られた。図9からわかるように、年齢が低い層は適応前の明瞭度が低く、しかも適応効果も大きい。図8と図9の結果から適応前の明瞭度が比較的低い話者ほど適応効果が大きいがわかる。明瞭度の低い音声であるが故に適応を必要とするとも言える。

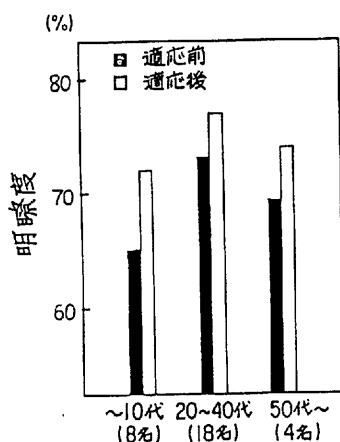


図9 話者の年齢と適応効果

5. 適応の消滅と音韻認識

5.1. 消滅特性を求める実験条件

これまでに、同一話者の音声を連続して聴くことによってその話者に対して適応し、音韻の認識率が上昇することを明らかにした。ここでは、ある話者に適

応した状態から、他の話者の音声を受聴することによって、元の話者への適応が失われる過程、即ち適応の消滅特性について検討した。

図10に適応の消滅特性を求める4種類の実験条件を示す。例えば条件aの場合には、音節の呈示位置1～5までは同一話者の音節が連続し、次に6の位置で異話者の音節が呈示された後、7の位置で再び元の話者の音節が呈示される。他のb、c、dの条件の場合も同様に、1から5と、それぞれ9、11、13の位置には同一話者の音節が、他の呈示位置には異話者の音節がランダムに呈示される。15個の音節の呈示が終ると、今度は話者を変えて再び1から同様の実験を繰り返す。

5.2.2. 適応の消滅特性

図11は、4つの実験条件による実験結果を同一図面上にプロットしたものである。4章に示した結果と同様に、5音節までは完全に適応がなされており、呈示位置1～5までが適応過程、5以降15までが消滅過程である。適応過程における結果が図5の結果と一致していないのは、実験回数と枠組が異なるためである。適応状態に達したのち、異話者の音節が呈示されると、受聴者の短期記憶に捕らえられている話者（声質）特徴に、それ以前の特徴と異なる新たな情報加わる。そのため元の話者について記憶されている情報が新しい情報によって消されていき、適応の消滅過程が生ずると考えられる。図中呈示位置9以降はこのことが示されており、適応状態が適応前の状態へと変化していくことがわかる。

ここで条件aの結果は、呈示位置7のところで元の話者の音節が呈示されても適応前の値に復帰せず、かなり明瞭度が低下することを示している。一方、条

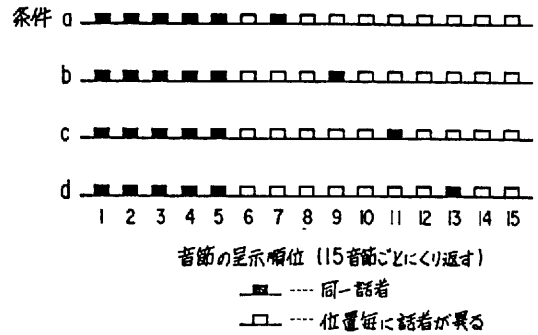


図10 適応の消滅特性を求める実験条件

件bの結果は、適応したあとで6、7、8と異話者の音節を受聴してから、9の位置で元の話者の音節を受聴したときの応答は、適応状態の値を保持しているという現象を示している。

この、一旦明瞭度が低下して次に回復するという状態はどのような聴覚情報処理が行われているのであろうか。この背景は必ずしも明確ではないが、一つ想像できることは、個人差への適応にもとづく音韻判断のstrategyが、6の位置の異話者の音節で一時的に乱されて、条件1の全く話者が異なっている場合と同じ状況が擬似的につくられてしまうのではなかろうか。2～3名の異話者の音節の受聴によって、その一時的な乱れが回復して、適応した元の状態に戻ることが考えられる。あるいは、6の位置で一時的に混乱状態に陥ると、次に元の話者の音声に到達したとき、7の位置では前の情報の検索に要する時間が音韻の認識に利用するのに間に合わないということも考えられる。これらの原因を解明するためには更に検討を進めなければならない。

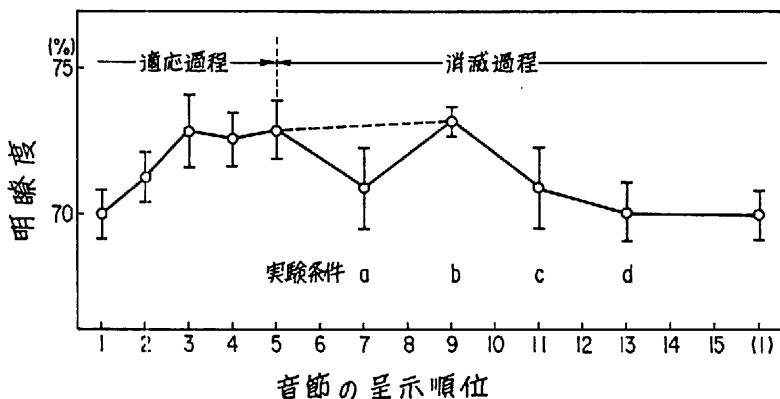


図11 話者に対する適応とその消滅過程

6. まとめ

以上に示したように人間が音声を受聴する際の音韻認識能力のうち、話者に対する適応能力について実験的検討を行った。対象とする音声を日本語の最も基本単位である音節に限定し、多数話者の発声した音節を種々の順序で呈示して、明瞭度の観点から検討を行った。音声の聴き取りは明瞭度試験に慣れている通話試験員によって行った。その結果、次のことが明らかとなった。

- (1) 話者に対する適応能力の存在が確認された。
- (2) その適応は4～5音節程度の音声の受聴で達成され、話者の個人性に対する知識が蓄積される。
- (3) 適応は同種の音素が呈示されたか否かには依存しない。
- (4) 音素や話者によって適応効果に差がある。子音別に見ると、/t,k,d,g/の破裂音などで効果が大きく、/w,j/などの半母音、弾音/r/、/k,j/を除く拗音などの効果は小さい。後続母音別には、/i/の場合に効果が大きく、/a/の場合に小さい。話者別には、年齢の低い話者など適応前の明瞭度が低い話者ほど適応効果が大きい傾向がある。
- (5) 同一話者の音節が連続して呈示されて十分適応したあと、異話者の音節を一つ聴くと、その次の音節の話者が前と同じであっても、明瞭度が一時的に低下する現象が生ずる。これは、個人性への適応にもとづく音韻判断のstrategyが、一時的に乱れるためと考えられる。
- (6) ある話者に適応された状態は6～7音節の異話者の音声の受聴で失われる。

今後は、音節の呈示間隔との関係、音声の物理的パラメータとの関係について検討する予定である。個人性への適応の過程において、音声の個人性に関する典型的な特徴(声質のクラス)が、あらかじめ多数記憶されていてそれが検索されるのか、話者が変わる毎に音韻体系の新たな学習がなされるのか、これらの巧みな組み合わせによるのか、この具体的メカニズムを明らかにすることは、極めて興味ある課題である。ここで実験的に明らかにしたように、人間による声の個人差への適応は非常に早く、効率が良い。このメカニズムの解明は、音声認識や音声合成の進展に重要な示唆を与え得ると考えられる。

〔謝辞〕

本検討の際、貴重な御意見、御討論をいただいた第四研究室・算室長、情報通信基礎研究部・河原調査員に感謝する。また、計算機の使用にあたって協力していただいた研究室員、被験者となられた通話試験員に感謝する。

〔参考文献〕

- (1) W. Strange and R.R. Verbrugge: "Consonant environment specifies vowel identity", J. Acoust. Soc. Amer., 60, 1, pp. 213-224 (1976)
- (2) M.J. Macchi: "Identification of vowels spoken in isolation versus vowels spoken in consonantal context", J. Acoust. Soc. Amer., 68, 6, pp. 1636-1642 (1980)
- (3) P.F. Assmann, T.M. Nearey and J.T. Hogan: "Vowel identification: Orthographic, perceptual, and acoustic aspects", J. Acoust. Soc. Amer., 71, 4, pp. 975-989 (1982)
- (4) 斎藤: "明りょう度試験法と明りょう度値の評価に関する考察", 研実報, 8, 6, pp. 845-853 (昭34)
- (5) 加藤、河原: "音声知覚における話者への適応性の検討", 音学聴覚研資, H-84-3 (昭59-01)
- (6) 加藤、河原: "音節認識における話者の適応能力の支配要因", 音学講論, 2-4-18 (昭59-03)
- (7) 斎藤、松田: "音韻識別における付加情報について", 音学講論, 2-2-17 (昭37-05)
- (8) 古井: "音声知覚におけるスペクトル変化情報の役割", 音学聴覚研資, H-85- (昭60-01)
- (9) K.N. Stevens, C.E. Williams, J.R. Carbonell and B. Woods: "Speaker authentication and identification: A comparison of spectrographic and auditory presentations of speech material", J. Acoust. Soc. Amer., 44, 6, pp. 1596-1607 (1968)