

論文 / 著書情報
Article / Book Information

論題(和文)	連続音声中の音節による個人性知覚
Title(English)	
著者(和文)	松井 知子, アーウィン・ポーラック, 古井 貞熙
Authors(English)	SADAOKI FURUI
出典(和文)	日本音響学会1993年秋季講演論文集, , , pp. 379-380
Citation(English)	, , , pp. 379-380
発行日 / Pub. date	1993, 10

連続音声中の音節による個人性知覚*

◎ 松井知子† アーウィン・ポーラック‡ 古井貞熙†
 (†NTT ヒューマンインタフェース研究所 ‡ミシガン大学)

1 まえがき

音声波に含まれる個人性情報は、音韻性情報と深く関わっているが、その関係はまだ完全には明らかにされていない。これまで、人間/コンピュータによる話者認識において有効な音韻の種類については、1960年代から80年代にかけて数多くの報告がなされている[1]。それらには、鼻音や母音は有効性が高く、破裂音や摩擦音は低いというような結果が示されている。しかし、いずれも概括的な知見が多く、有効な音韻の種類に関する個人差については検討されていない。一方、日常会話において、/a/ が /e/ に近いとか、/h/ が /s/ になるなど、ある種の音韻が特徴的な話者に出会うことがある。そして、そのような話者を識別する時には、自然にその特徴的な音韻を手がかりにしている。従って、話者認識に有効な音韻の種類についても、各話者による差異があると思われる。本論文では、連続音声中の CVC (consonant-vowel-consonant) 音節を切り出して、人間による話者識別実験を行うことにより、話者認識に有効な母音の種類に個人差があることを明らかにする。

2 実験

本実験では、データ量や実験時間を考慮し、5 母音だけに注目して、その話者認識に対する有効性を調べた。使用したデータは、男8名が発声した文章データから切り出した。母音部だけを切り出した場合、その平均時間長は 90.5 ms と短く、母音部だけを聴いて話者を判定することは非常に困難である。そこで、CVC 音節単位で切り出して、それを被験者に3回づつ提示し、被験者はその話者を判定して、その結果を回答用紙に記入するという方法で実験を行った。被験者はあらかじめ話者のセットについて知らされており、判定に自信がない場合も、必ずどの話者が記入するように指示されている。なお、各被験者はどの話者の声もよく知っている。表1に実験条件を示す。

以下、本実験を計画するにあたり、特に考慮した点について述べる。

基本周波数の影響

基本周波数は個人性知覚に大きく寄与することが知られている[2]。使用するデータに、基本周波数が非常に異なる話者が含まれていると、被験者はその違いを手がかりに話者を判定してしまい、うまく母音間の差異が測定できない危険性がある。そこで、本実験では平均基

表 1: 実験条件

話者	男 8 名
被験者	11 名 (男 5 名、女 6 名)
レシーバ	STAX SRM-1 (片耳受聴)
受聴レベル	聞きやすいレベル
データ	720CVC 音節 (= 5 母音 × 8 話者 × 18) 6kHz 帯域 平均時間長 167.8 ms

本周波数が近い話者 (平均 165 ~ 175 Hz) を選択して使用した。

時間長の影響

連続音声データ中には、発声のなまけから、非常に時間長の短い母音が存在する。また、CVC 音節単位で切り出した時に、子音部の時間長があまり長いと、被験者は子音部に影響され、母音による差異をうまく測定できない。本実験では、母音部が 50 ms 以上の CVC 音節だけを使用し、また 30 ms 以上の子音部については、窓をかけてカットした。

文脈の影響

表1にあるように、CVC 音節は各話者、各母音については 18 個づつ均等に用意した。母音部が同じでも、子音部が異なれば、文脈の影響で話者判定のしやすさが変わってくると思われる。全子音の組み合わせについて、均等に CVC 音節を用意することが望ましいが、データ量が膨大になり、実験を行うことが困難になる。本実験では、子音の組み合わせが話者によって等しくなることだけを考慮して、各話者同じ 10 文章から、CVC 音節を切り出した。従って、これに含まれる子音の種類に関しては偏りがある。

3 結果

表2は、全被験者の話者識別率を平均した結果である。この結果をもとに、次の二つの分析を行った。

第一に、話者または母音の水準間で識別率に差があるかどうかを分散分析によって調べた。母音についての分散比 (自由度 7, 28) は 6.1 で、有意水準 0.001 の F 値 4.9 より大きく、母音間には差は認められなかった。話者についての分散比 (自由度 4, 28) は 1.6 で、有意

*Perception of voice individuality using syllables in continuous speech.

By Tomoko Matsui†, Irwin Pollack‡, and Sadaoki Furui† (†NTT Human Interface Laboratories, ‡Michigan University)

表 2: 平均話者識別率 (%)

話者	/a/	/e/	/i/	/o/	/u/	話者平均
S1	66.7	73.7	44.4	63.1	51.5	59.9
S2	56.1	20.2	33.8	49.5	38.4	39.6
S3	59.6	66.7	80.8	53.5	44.4	61.0
S4	40.9	44.4	50.0	58.6	67.2	52.2
S5	73.2	81.3	69.2	77.3	72.7	74.8
S6	93.9	84.3	42.4	78.3	93.4	78.5
S7	54.6	40.9	33.8	41.4	27.8	39.7
S8	59.6	48.5	37.4	77.8	48.5	54.3
音韻平均	63.1	57.5	49.0	62.4	55.5	57.5

表 3: 話者識別に最も有効な母音と有効でない母音

話者	有効	有効でない
S1	/e/	/i/
S2	/a/	/e/
S3	/i/	/u/
S4	/u/	/a/
S5	(母音間に差なし)	
S6	/a/, /u/	/i/
S7	/a/	/u/
S8	/o/	/i/

水準 0.05 の F 値 2.7 より小さく、話者間には差があることがわかった。

第二に、各話者ごとに、各母音の識別率に偏りがあるかどうか調べるために、 χ^2 適合度検定を行った。各話者の χ^2 値 (自由度 4) は、

S1: 18.4 S2: 39.0 S3: 24.5 S4: 17.3

S5: 2.3 S6: 45.3 S7: 20.0 S8: 34.0

であり、S5 を除くと、有意水準 0.05 の χ^2 値 9.5 よりも大きい。従って、各話者ごとの各母音の識別率には偏りがある。

上記の二つの分析結果は、話者識別に有効な母音が話者によって異なることを示している。そこで、実際に各話者ごとに、その話者の識別に最も有効な母音、有効でない母音について調べた。表 3 にその結果を示す。

4 考察

4.1 逆転データによる実験

CVC 音節からも、話し方のくせ、韻律情報に含まれる個性を知覚することができる。韻律情報は、本実験の結果にも何らかの影響を及ぼしている可能性がある。そこで、データの時間軸を逆転させ、韻律情報の影響を取り除くことを考える。表 4 に、この場合の各被験者の話者識別率を平均した結果を示す。

これをもとに、分散分析、 χ^2 適合度検定を行ったが、順方向データによる結果と全く等しい結果が得られた。順方向、逆方向に関わらず、話者識別に有効な母音が特定できることがわかった。

以上の結果から、人間は CVC 音節を聞いて話者を判定する時に、その音節の静的なスペクトル情報に含まれ

る個性を主に利用していることがわかった。

4.2 データ時間長と識別率との関係

データ時間長と識別率との関係について調べた。本実験で使用した CVC 音節を、その母音部によって分類した時の CVC 音節の平均時間長 [ms] は、
/a/:171.1 /e/:182.6 /i/:150.6 /o/:168.2 /u/:166.4
であった。各話者ごとに、各母音の識別率とデータ時間長との相関を調べた結果、それぞれの相関係数は、

S1: 1.0 S2: -0.1 S3: -0.3 S4: -0.5

S5: 0.9 S6: 0.5 S7: 0.8 S8: 0.3

であった。無相関の仮説に対する有意水準 0.1 の相関係数 0.81 よりも大きいのは、S1 と S5 だけで、他の 6 話者については各母音の識別率とデータ時間長に相関が認められなかった。

5 まとめ

本論文では、連続音声の中の CVC 音節を切り出して、人間による話者識別実験を行い、話者認識に有効な母音が話者によって異なることを示した。また、人間は CVC 音節を聞いて話者を判定する時に、その音節の静的なスペクトル情報に含まれる個性情報を主に利用していることがわかった。

今後は、本実験のデータを用いてコンピュータによる話者識別実験を行い、その結果と本論文の結果とを比較する予定である。

謝辞

実験にご協力頂いた、ヒューマンインタフェース研究所音声部の栗原研究員、小口研究員、および古井特別研究室の方々に深く感謝致します。

参考文献

- [1] J.P. Eatock and J.S.D. Mason, "Phoneme performance in speaker recognition," *Proc. ICSLP, Banff*, pp. 11-1411-1414 (1992)
- [2] 伊藤憲三、齊藤収三、"音声の音響的特徴パラメータが個性の知覚に及ぼす影響," 信学論、J65-A、1、pp.101-108 (1982)

表 4: 逆転データによる平均話者識別率 (%)

話者	/a/	/e/	/i/	/o/	/u/	話者平均
S1	62.6	78.3	54.6	60.1	52.0	61.5
S2	57.6	20.7	30.8	42.4	34.3	37.2
S3	56.1	67.7	84.3	52.5	59.6	64.0
S4	40.9	26.3	53.0	56.1	60.1	47.3
S5	75.3	71.7	61.1	66.7	66.2	68.2
S6	77.8	79.3	44.4	79.8	83.8	73.0
S7	61.6	39.4	40.4	46.0	30.8	43.6
S8	49.0	36.4	51.0	79.8	55.6	54.3
音韻平均	60.1	52.5	52.5	60.4	55.3	56.2