聴取による話者識別における言語の影響
―日本語母語話者が発話した日本語・英語の場合―*

☆井上崚河（上智大），網野加苗（科警研），荒井隆行（上智大）

1　はじめに

聴取による話者識別において，言語情報がどのように影響を及ぼすのかという研究はこれまでにも行われてきた．例えば Winters ら [1]では，バイリンガル話者を識別する際，言語に依存しない情報（音声の個人性）と言語に依存する情報（言語情報）の両方を利用して，識別が可能になることが報告されている．さらに Wester ら [2]では，2 つの刺激音が同一言語の場合の話者識別では，聴取者は音声の個人性にのみ焦点を合わせることができるため，音声信号内の言語情報による影響を受けないことが示された。一方，刺激音の言語が異なる場合においては，聴取者は，慣れ親しんだ言語の音声については音声の個人性と言語情報の両方を利用できるものの，聴取者にとっての外国語の音声では個人性のみを利用することが報告されている．また Mok ら [3]では，聴取者の言語の習熟度は話者識別における自信度と相関がみられること，話者識別の精度は刺激音の言語が同じである場合の方が，言語が異なる場合よりも高くなることが述べられている．どの先行研究においても，話者が母語を話す場合の識別精度は，非母語を話す場合の識別精度よりも高くなったと報告されている．

しかしながら，複数言語の一方が日本語である場合の話者識別に関する研究は行われていない．日本人には英語を聞き取ること，そして話すことに苦手意識を持っている人が多いということをよく耳にする．日本語を母語とする聴取者が，日本語と英語を話す日本語母語話者を識別する場合，話者及び聴取者にとっての母語である日本語の方が識別精度が高くなる可能性と，先の非母語である英語に対する苦手意識による発音の不安定さを要因として英語の方が識別精度が高くなる可能性の両方が考えられる．

これを解明するため本研究では，日本語を母語とする話者と聴取者を対象とし，日本語音声と英語音声を用いた場合の話者識別精度の違いを調べた．また，幼少期（12 歳以下）における海外滞在期間が 2 年未満の日本語母語話者（以後 N），幼少期（12 歳以下）における海外（英語圏）の小学校（日本人学校を除く）在籍期間が 2 年以上の日本語母語話者（以後 R）に協力を依頼し，英語の流暢さの違いによる識別精度への影響を調べた．

2　録音

2.1 話者

録音には，N と R からそれぞれ男性 4 名，計 8 名（平均年齢 20.5 歳）が参加した．また，参加者は自己申告で全員健聴であった．

2.2 単語

話者に読み上げてもらう単語は，日本語は天野ら [4]から，英語は西出ら [5]から，親密度が 90%（[4] では文字音声単語親密度 6.3/7, [5] では評点 4.5/5）以上の，3 音節で 7 音素から成る名詞「今月」「自転車」「本日」「親切」，‘company’‘average’‘negative’‘butterfly’の 8 語を選定した．

2.3 手順

録音はすべて防音室（上智大学荒井研究室）で行った．すべての音声は，マイクロフォン（ECM-MS957，SONY）とオーディオインターフェース（USB AudioCapture UA-25EX, Roland）を介して，コンピュータ（Surface Pro, Microsoft）上のソフトウェア（Audacity [6]）を用いて録音した．標本化周波数は 48 kHz，量子化精度は 16 bit とした．単語はすべてキャリア文（日本語は「これを～といいます」，英語は‘Repeat ～, please’）に入れた．それらが記載されたスライドをランダムに提示し，読み上げてもらった．この試行を 4 回繰り返

* The effects of language on speaker identification -In case of Japanese and English spoken by Japanese native speakers-, by Ryoga INOUE (Sophia Univ.), Kanae AMINO (NRIPS) and Takayuki ARAI (Sophia Univ.).

した．すなわち，話者 1 名あたり 32 文（単語 8 種類，試行回数 4 回）を録音し，合計で 256 文（話者 8 名，1 名あたり 32 文）の音声が得られた．数回の練習と任意の休憩を含め，録音の所要時間は約 20 分であった．

3　聴取実験

3.1 聴取者

聴取実験には，海外滞在期間が 2 年未満の日本語母語話者で，録音に参加した話者との関わりがない，男性 18 名，女性 7 名，合計 25 名（平均 21.7 歳）が参加した．また，参加者は自己申告で全員健聴であった．

3.2 刺激音

ソフトウェア Praat [7]を用いて，録音した音声の中から不明瞭な発話を除外した. 次に，残ったすべての発話に対して，キャリア文全体と単語部分のみの平均基本周波数を調べた. 以後，それぞれ F0c，F0 とする．全話者の中から F0c の値が近い者を, R, N それぞれ 2 名ずつ選んだ．また，4 名の発話において，8 語の中から, F0 の値が近い単語を日英 2 語ずつ選んだ．最終的に，その 4 語（「自転車」「本日」‘company’‘butterfly’）について，1 話者につき 2 トークンずつ選び，単語部分のみを刺激音として聴取実験に用いた．そのため，選定した刺激音は 32 個（単語 4 種類，話者 4 名，各単語 2 個ずつ）となった．

なお，話者個々の声量の差や，マイクロフォンとの距離の差によって生じる刺激音の音量差を解消するため，あらかじめ Praat で作成したプログラムを用いて振幅値を 80 dB に調整することで一定にした．

3.3 手順

聴取実験はすべて防音室（上智大学荒井研究室）または無響室（上智大学音声研究室）にて行った．刺激音は，ソフトウェア MATLAB [8]で作成した実験用 GUI を用いて AX 法にて提示した．2 つの刺激音を A，X の順に提示し，A の話者と X の話者が同一の話者であるか，異なる話者であるかを強制選択させた．順序効果を抑えるため，単語はすべての組み合わせでランダムに提示した. また，同一話者で同一単語同士の組み合わせに関しては，異なるトークンを用いた．

実験参加者は，コンピュータからオーディオインターフェースを介し，ヘッドフォン（HDA200，Sennheiser）より刺激音を聴取した．各刺激音は 1 度だけ再生し，正誤のフィードバックは行わなかった．刺激音の組み合わせは 256 通り（話者の組み合わせ 16 通り，単語の組み合わせ 16 通り）であり，得られる正答率のデータの信頼度を高めるために，繰り返し回数を 2 回とした．また，任意で 4 回の休憩（1 回あたり 5 分以内）をはさみ，実験所要時間は全体で約 50 分であった。

4　結果

以下では, 刺激音 A と X の話者の組み合わせ条件（以後，話者条件）NN，NR，RN，RR および言語の組み合わせ条件（以後，言語条件）JJ（日本語，日本語），JE（日本語，英語），EJ（英語，日本語），EE（英語，英語）について，条件ごとに考察する．また，各条件における平均正答率を比較した（図 1-4，表 1）．

実験の結果，どの話者条件下でも，JJ よりも EE における正答率（以後，識別精度とする）の方が高くなった．また，JE や EJ のように刺激音の言語が異なる場合，識別精度が低くなった．さらにこの JE，EJ について，NN，NR，RR の条件下では EJ における識別精度の方が JE より高くなったが, RN では JE における識別精度の方が高くなった．

一方，それぞれの言語条件における識別精度を比較すると，JJ では，NN，RR における識別精度が高くなり，NR，RN において低くなっている．JJ 以外の条件下では，話者条件 NR, RN での識別精度の方が, 話者条件 NN, RR での識別精度よりも高くなっていることがわかる．

5　考察

結果から，刺激音がともに日本語の場合よりも，ともに英語の場合の識別精度の方が高くなったことがわかる．発話における訛りが個人性に関係することは峯松 [9] においても指摘されている．このことから，日本語母語話者が非母語である英語を話す際，母語である日本語ほど流暢に発話することができないため，発音の訛りにおける個人性の表れやすさに差が生じ，その結果識別精度にこのような傾向が表れた可能性が考えられる．

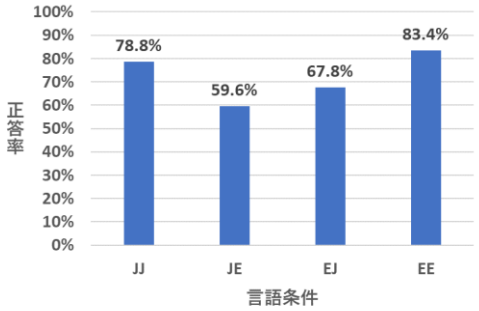

図 1　話者条件 NN における識別精度

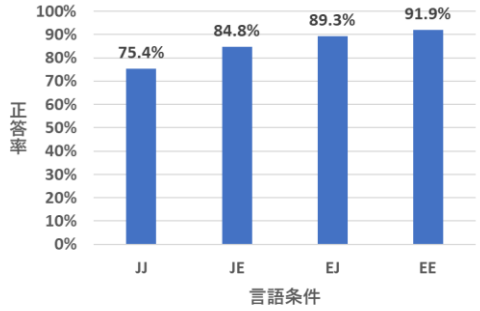

図 2　話者条件 NR における識別精度

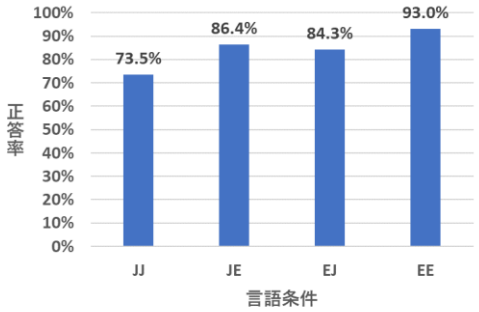

図 3　話者条件 RN における識別精度

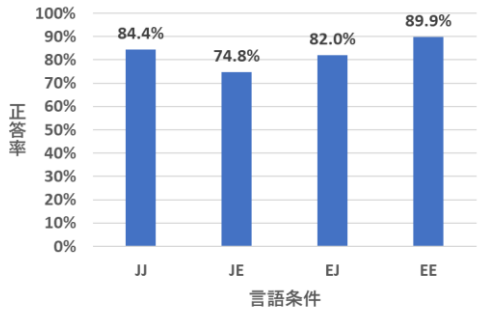

図 4　話者条件 RR における識別精度

表 1. 各条件での識別における標準偏差

	NN	NR	RN	RR
JJ	9.7%	16.1%	18.1%	8.6%
JE	8.4%	15.3%	12.5%	14.5%
EJ	12.2%	11.4%	13.0%	11.3%
EE	11.7%	11.4%	9.0%	9.2%

また，NN や RR で刺激音の言語が異なる条件下において識別精度が低くなったのは，Mok ら [3]で報告されていた，「刺激音が同一言語の場合の識別精度が，言語が異なる場合の識別精度より高くなる」ことを支持する結果である．しかし，NR や RN において JE や EJ は，EE よりは識別精度が低いものの，JJ よりは高くなっていたことから，話者が N と R の組み合わせの場合，刺激音の言語が異なる条件下であっても，ある程度高い識別精度を保つことができると考えられる．先にも述べた通り，話者条件 NR と RN では NN や RR に比べ識別精度は高くなっている．この NR，RN において識別精度が高くなった理由としては，聴取者が，話者の非母語の発音やアクセントに表れる流暢性や訛りの違いを，個人性として利用した可能性が考えられる．

一方，JJ において NN と RR における識別精度が高くなり，NR と RN において低くなったことがわかるが，これは，話者の言語背景の類似性によって識別精度が高くなった可能性を示唆している．

聴取実験後のフォローアップでは，「2 つの単語がともに日本語のパターンは難しかった.」，「2 つの単語が両方とも英語のパターンは比較的わかりやすかった.」「日本語と英語が混ざると難易度が上がった.」などの意見があり，これらは本研究にて得られた結果を裏付けているといえる．また，手順にて述べたように本実験の試行回数は 2 回であり，全体的には 2 回目の試行における識別精度が 1 回目よりも高くなった．しかし，なかには 1 回目は正解しているが，2 回目で不正解というケースも見られた．まず識別精度が全体的に上がったのは，学習効果が一番の要因である考えられる．一方，2 回目で不正解となった要因としては，実験による疲労が原因となったのではないかと考えられる．

本研究における聴取実験で提示した刺激音の組み合わせは，同一話者が全体の 1/4，異なる話者が 3/4 であった．これは，例えば参加者が全回答を「同じ」とした場合よりも，「異なる」とした場合の識別精度がより高くなることになる．また，聴取者が実験に参加している最中，「異なる」を選択する試行が多いため，参加者の回答が「異なる」に偏った可能性や，「同じ」と「異なる」の試行数の違いが

バイアスとなり識別精度に影響を及ぼした可能性も考えられる．聴取者が「話者が同じ」と答える回数と「話者が異なる」と答える回数を，それぞれ全体の 1/2 と等しくすることで，これらの影響がない状態にした実験が理想である．これは今回の研究における反省点であり，今後はこのことを考慮したうえで，実験内容を改善したい．

6　おわりに

本研究では，複数言語による話者識別の聴取実験に初めて日本語を取り入れた．その結果として，話者の条件に関係なく，刺激音が日本語同士の場合よりも英語同士の場合における識別精度の方が高くなるという傾向がみられた．このことから，日本語を母語とする聴取者が，日本語と英語を話す日本語母語話者を識別する場合，非母語である英語同士において識別精度が高くなることが示された．また，話者条件が NN，RR のように一致しているとき，刺激音の言語が同じ場合の識別精度は，刺激音の言語が異なる場合の識別精度よりも高くなり，Mok ら [3]を支持する結果となった．

しかしながら，本研究では，話者としても聴取者としても，日本語母語話者のみを対象とした．そのため，英語母語話者を話者，聴取者の対象に加えることで，より厳密に日本語と英語の 2 言語における，母語非母語と識別精度の関係性を具体的な結果として得ることができるだろう．さらに，聴取時の自信度も併せて調べることによって，Mok ら [3] によって報告された，言語習熟度との関連も明らかにしたい．

本研究で使用した刺激音の話者について，R による英語はネイティブに近い発音であり，俗に言う「日本語英語」は含まれなかった．一方，N による英語には， ネイティブに近い発音は含まれなかった．今後は，ネイティブスピーカーによる訛りの度合い（accentedness）の評価をすることで，より細かく分析をしたい．

また，今回の実験の参加者は，録音にて 8 名（うち聴取実験に音声使用したのは 4 名），聴取実験にて 25 名と，必ずしも多いとは言えない．そのため，刺激音を選定する際に比較した平均基本周波数に，同じ話者であっても単語によって差があるものや，ほかの単語と比べると標準偏差が若干高くなってしまった単語があった．今後，実験参加者数を増やすことで，より信頼度の高いデータが得られることが期待できる．

謝辞

本研究の実施にあたり，上智大学倫理委員会の承認を得た．また，実験に参加してくださいました皆様に，心より御礼申し上げます．

参考文献

[1] Winters, et al., Acoust. Sci. Tech., 123 (6), 4524-4538, 2008.

[2] Wester, Speech Commun., 54 (6), 781-790, 2012.

[3] Mok, et al., Int. Journal of Speech, Language & the Law, 22 (1), 55-77, 2015.

[4] 天野，近藤，*日本語の語彙特性-単語親密度-, NTT データベースシリーズ第 1 巻*, 三省堂, 1999.

[5] 西出，水本，*英単語 8000 語についての親密度測定の試み*, 都留文科大学大学院紀要, 13, 57-92, 2009.

[6] The Audacity Team, https://www.audacityteam.org/

[7] Boersma and Weenink, “Praat: doing phonetics by computer.” Retrieved from http://www.praat.org/

[8] The MathWorks, http://www.mathworks.com/

[9] 峯松，音響誌, 71 (9), 490-497, 2015.