言語的符号化の有無を要因とした記憶実験で使用する音声刺激の作成

井上, 晴菜

(出版者 / Publisher)
法政大学大学院
(雑誌名 / Journal or Publication Title)
大学院紀要 = Bulletin of graduate studies
(巻 / Volume)
87
(開始ページ / Start Page)
30
(終了ページ / End Page)
38
(発行年 / Year)
2021-10-31
(URL)
https://doi.org/10.15002/00024763

言語的符号化の有無を要因とした記憶実験で使用する音声刺激の作成[1][2]

人文科学研究科　心理学専攻
博士後期課程 1 年　井上 晴菜

21 世紀に入って以降，言語化（音声の特徴や印象を言葉で表す符号化方略）による話者同定の成績への影響が検討されている。しかし，結果が一貫しておらず，その原因は，音声ラインナップを構成する標的音声と妨害刺激の言語化した場合の類似性にあると思われる。先行研究では，標的音声との全体的な「聴いた印象」の類似度が低い妨害刺激を選定しているが，このとき，必ずしも標的音声との言語化情報の類似度も低いとは限らない。本研究では，標的音声との言語化情報の類似性を基に妨害刺激を選定する方法を提案した。具体的には，各音声に対していくつかの形容詞対を用いた評定を行い，音声刺激間で類似している音声刺激（高類似音声），あるいは類似していない音声刺激（低類似音声）の候補を挙げるために多次元尺度構成法とクラスター分析を行った。また同時に，その刺激の候補が十分に弁別可能なものであることを確認するために，弁別課題を行った。それにより，音声に関する幾らかの次元（例えば，「声の高さ」「声の太さ」）について言語化を行った類似度を指標として，「標的音声」と，その標的音声の言語化情報との類似度が高い「高類似音声」およびその類似度が低い「低類似音声」を選定した。その結果と考察を示す。

Key words: similarity, verbalized information, speaker identification

我々は通常，顔や声などの情報を活用し，人物を同定することができる。それでは，顔は見えないが声は聞こえる状況で，その声からその人物を同定（話者同定）することはできるのだろうか。以下，詳しく説明する。

話者同定（speaker identification）とは，特定の人物の音声が一人または二人以上の人物の音声と区別される処理のことである（Yarmey, 2007）。話者同定の典型的な手続きとしては，話者同定テスト時に生のまたは録音の音声を聴いた後に，音声の特徴や印象（例えば，声の高さや太さ）などを手がかりに，学習時に聴いた音声の話者と同一かどうかを判断する。この際，学習時とテスト時の音声刺激自体は同じではなく，異なるのが一般的であるという点で，音声再認とは区別される。音声再認の場合は，再認テスト時に録音の音声刺激を聴いた後に，学習時に聴いた音声刺激と同一かどうかを判断する。つまり，音声再認の課題遂行には，音声の特徴や印象だけでなく，背景雑音なども手がかりになり得る。そのため，音声再認は，話者同定では役に立たないような情報からでも再認が可能であるといえる。なお，これより先は，音声再認ではなく話者同定について論じる。

話者同定の課題遂行は，聴者にとって話者が，今までに音声を何回も聴いたことがある人か，または，今までに音声を一回は聴いたことがある人か，という話者の音声の聴取頻度に影響を受ける。これは，Yarmey, Yarmey, Yarmey, & Parliament（2001）の実験から説明できる。 Yarmey et al.（2001）は，68 人の音声刺激を使用し，その中から 4 人（各参加者にとって「よく知っている人」，「ある程度知っている人」，「音声を二，三回は聞いたことがある人」「音声を一回も聞いたことがない人」）の音声を提示した。その結果，一回も聞いたことがない人の音声から正しく「知らない人である」と報告できた確率（55%）および，二，三回は聞いたことがある人の音声から正しく他の人と識別ができた確率（49%）の間で差がなく，両者とも，よく知っている人の音声から正しく他の人と識別ができた確率（85%）より低かったという。このことから，よく知らない人の音声に対する話者同定は，よく知っている人の音声に対するそれよりかなり難しいと解釈できる。

Correspondence concerning this article should be sent to: Haruna Inoue, Psychology, Graduate School of Humanities, Hosei University, 2-17-1, Fujimi, Chiyoda-ku, Tokyo 102-8160, Japan. (E-mail: h.inoueee10@gmail.com)

[1] 本研究結果の一部は，日本心理学会第 83 回大会（2019）で発表された。

[2] 本論文は，令和元年度に法政大学大学院人文科学研究科へ提出した修士論文の一部を加筆・修正したものである。

よく知らない人の音声とよく知っている人の音声では，何が違うのだろうか。容易に思い至るのは，その音声に関する豊富な既有知識があるかどうかの違いだろう。よく知っている人の音声は，接する機会が多く，例えば，声の高さや太さといった音声の特徴や印象に関する多種多様な知識があると考えられる。一方，よく知らない人の音声は，それらの知識がないに等しいだろう。さらに，その音声を聴く機会が少ないため，その一回の機会で，話者同定ができる程度の情報を獲得する必要がある。その方略の一つに，音声の特徴や印象を言葉で表す言語的符号化（以下，言語化とする）があり，21 世紀に入って以降，話者同定の成績への影響が研究されている。しかし，音声の言語化による話者同定の成績への影響は一貫しておらず，促進効果と抑制効果という両極の効果が得られている。なお，後者のように，記銘材料に対して言語化を行うことが記憶成績を抑制する現象を言語隠蔽効果（verbal overshadowing effect；Schooler & Engstler-Schooler, 1990）という。

言語化の有無を要因とした実験

音声の言語化により，話者同定の成績に促進効果のパターンの結果を得た実験は少ないが，その一つに田中（2017）の実験がある。まず学習として，標的音声を提示した。次に，各群の条件に従って異なる教示を与えた。音声を言語化する群（言語化群）には，標的音声の特徴を思いつく限り説明するように指示した（自由記述）。一方，言語化しない群（統制群）には，一桁の足し算の 100 マス計算と，一桁の掛け算の 100 マス計算を解くように指示した。なお，各群の制限時間は 5 分間であった。その後，話者同定テストとして，標的音声を含む 6 名の音声ラインナップを提示した。ラインナップとは，（学習時に提示した）標的音声を含むまたは含まない複数の音声刺激を順に提示する方法である[3]。そして，全ての音声刺激を聴いた後に，「標的音声と同じ人物だと思う」1 人を選ぶように求めた。その後，話者同定テストの回答に対する確信度の評定を行ってもらい，実験を終了した。結果は，正同定率，すなわち標的音声を正しく選ぶ率が，統制群よりも言語化群のほうが高くなった。つまり，田中（2017）の実験では，言語化により話者同定の成績に促進効果のパターンがみられた。

これに対して，音声の言語化により，話者同定の成績に言語隠蔽効果のパターンの結果を得た実験に，Perfect, Hunt, & Harris（2002）と Vanags, Carroll, & Perfect（2005）の実験がある。それらの実験では，まず学習として，標的音声を提示した。次に，挿入課題を挟み，各群の条件に従って異なる教示を与えた。音声を言語化する言語化群には，標的音声の特徴をできる限り詳細に説明するように指示した（自由記述）。一方，言語化しない統制群には，Perfect et al.（2002）では何もせず待機するように指示し，Vanags et al.（2005）では 9 つのアルファベットから単語を生成するように指示した。なお，各群の制限時間は 5 分間であった。その後，話者同定テストとして，標的音声を含む 6 名の音声ラインナップを提示した。そして，全ての音声刺激を聴いた後に，Perfect et al.（2002）では標的音声と同じ人物だと思う 1 人を選ぶように求め，Vanags et al.（2005）では標的人物がいると思った場合はその 1 人を選び，いないと思った場合は「いない」と回答するように求めた。その後，話者同定テストの回答に対する確信度の評定を行ってもらい，実験を終了した。いずれの研究においても，正同定率，すなわち標的音声を正しく選ぶ率が，統制群よりも言語化群のほうが低くなった。つまり，Perfect et al.（2002）と Vanags et al.（2005）の実験では，言語化により話者同定の成績に言語隠蔽効果のパターンがみられた。

なぜ，音声の言語化により話者同定の成績に促進効果と抑制効果という両極の効果が得られたのだろうか。それには，音声ラインナップを構成する標的音声とそれ以外の音声刺激（妨害刺激）の類似性が影響した可能性がある。すなわち，音声刺激の選定方法に問題があるのではないか。

言語化の有無を要因とした実験における音声刺激の選定方法

田中（2017），Perfect et al.（2002）および Vanags et al.（2005）では，標的音声以外の妨害刺激はどのように選定されたのか。田中（2017）と Perfect et al.（2002）の研究では選定方法に関する記述がなかったが，Vanags et al.（2005）では聴覚提示された音声刺激の全体的な「聴いた印象」の類似性の評定を基に，互いに

[3] 標的音声または標的音声以外の音声刺激（妨害刺激）のどちらか一つの音声刺激を単独で提示するショウアップとは区別する。

似ていない音声刺激を選定したという。この選定方法は，言語化を要因とする研究以外の，音声刺激を用いた研究でも採用されていた（e.g., Olsson, Juslin, & Winman, 1998; Orchard & Yarmey,1995; McDougall, Nolan, & Hudson, 2015; Yarmey, 2001）。しかし，必ずしも，「聴いた印象」の類似度が低くても，言語化をした場合の類似度も低いとは限らない。その傍証となる実験に北神（2000）がある。

北神（2000）は，視覚刺激を用い，標的図形との形態は類似しており，標的図形の言語的なラベル，すなわち標的図形の形態を言語化した情報との類似性が異なる 2 種類の妨害刺激を用い，言語化が記憶成績に促進効果または抑制効果を与えることを再現した。2 種類の妨害刺激のうち一つは，標的図形と，形態は類似しているが，言語的なラベルが適合していない図形，すなわち標的図形を言語化した情報との一致度が低い図形であった。もう一つは，標的図形と，形態も類似しており，言語的なラベルも適合している図形，すなわち標的図形を言語化した情報との一致度が高い図形であった。なお，標的図形のラベルは，参加者が自由記述で生成したものではなく，実験者があらかじめ用意したものであった。その結果，記憶テスト時に標的図形とともに，標的図形のラベルとの適合度が低い図形が提示された場合は，標的図形のラベルがある条件の方がない条件よりも標的図形の選択率が高かった（促進効果）。一方，テスト時に標的図形とともに，標的図形のラベルとの適合度が高い図形が提示された場合は，標的図形のラベルがある条件の方がない条件よりも標的図形の選択率が低かった（抑制効果）。この結果から言えることは，言語化した情報により妨害刺激と区別ができる場合は，記憶テストにおいて標的刺激を標的刺激だと正しく選択しやすくなるため，標的刺激の再認成績に促進効果をもたらし，逆に，言語化した情報がその妨害刺激にも当てはまってしまい，言語化された情報により妨害刺激と区別ができない場合は，記憶テストにおいて妨害刺激を標的刺激だと誤って選択しやすくなるため，標的刺激の再認成績に抑制効果をもたらすと解釈できる。

このように，必ずしも，標的音声と妨害刺激の「聴いた感じ」の類似度が低（または高）くても，言語化をした場合の類似度も同様に低い（または高い）とは限らないと言える。すなわち，（促進効果のパターンが得られた）田中（2017）は，言語化をした場合の標的音声との類似度が低い妨害刺激が選定されており，一方で，（言語隠蔽効果のパターンが得られた）Perfect et al.（2002）と Vanags et al.（2005）は，言語化をした場合の標的音声との類似度が高い妨害刺激が選定されていた可能性がある。すなわち，音声の言語化により話者同定の成績に促進効果と抑制効果という両極の効果が得られた原因は，標的音声に対する言語化の影響というより，標的音声と妨害刺激の言語化をした場合の類似性の影響にあるのではないか。具体的には，標的音声を言語化した情報により妨害刺激と区別ができる場合は，話者同定テストにおいて標的音声を標的音声だと正しく選択しやすくなるため，田中（2017）のように標的音声の話者同定の成績に促進効果のパターンをもたらし，逆に，標的音声を言語化した情報がその妨害刺激にも当てはまってしまい，言語化した情報により妨害刺激と区別ができない場合は，話者同定テストにおいて妨害刺激を標的音声だと誤って選択しやすくなるため，Perfect et al.（2002）と Vanags et al.（2005）のように標的音声の話者同定の成績に言語隠蔽効果のパターンをもたらしたのではないかと考えられる。しかし，田中（2017），Perfect et al.（2002）および Vanags et al.（2005）には，標的音声と妨害刺激の言語化した場合の類似性については明記されていないため，あくまで推測である。

以上のことから，言語化の有無を要因とした研究では，言語化の影響を明確にするためにも，類似性は「言語化をしていない場合（ここでは，全体的な「聴いた印象」の類似性）」と「言語化をした場合」を分離した上で，言語化をした場合の類似性をもとに妨害刺激を選定する必要があると思われる。そこで，本研究では，言語化をした場合の類似性をもとに音声刺激を選定する方法を提案する。

音声刺激を言語化した場合の類似性による妨害刺激の選定方法

まず，音声刺激の言語化の対象について考える。北神（2000）では，視覚刺激のラベル付けの対象は，記憶テストの課題遂行に有効な手がかりである「形態」であった。そして，視覚情報が類似していても，言語情報が類似していない妨害刺激との比較では標的刺激が選ばれ，視覚情報が類似しており，言語情報も類似している妨害刺激との比較では妨害刺激が選ばれた。すなわち，音声刺激において，話者同定の課題遂行に有効な手がかりとなるのは，「音声の特徴や印象」であるため，それを言語化の対象とするのが妥当であると思われ

る。すなわち，ある音声刺激に対して「高い声」「太い声」などのラベルを付けることになる。本研究では，このように，音声の特徴や印象を言葉で表すことで生み出された言語的な情報を，声質の言語化情報（verbalized information）と呼ぶ。

次に，本研究では，標的音声との言語化情報の類似性を基に，類似している妨害刺激と類似していない妨害刺激を次の通り定義する。標的音声との言語化情報の類似度が低い音声とは，例えば，「声の太さ」の次元では標的音声の言語化情報（例：細い声）と一致しない「太い声」と言語化される音声と定義する。反対に，類似度が高い音声とは，標的音声の言語化情報（例：細い声）と一致する「細い声」と言語化される音声と定義する。

従って，本研究では，類似度が低い妨害刺激には，言語化情報が標的音声と一致していない音声刺激（以降，低類似音声）を選定し，類似度が高い妨害刺激には，標的音声との弁別自体は十分可能だが，言語化情報が標的音声と一致している音声刺激（以降，高類似音声）を選定する。すなわち，Vanags et al.（2005）と同様に，標的音声と妨害刺激は「聴いた印象」が互いに類似していない音声刺激となる。なお，妨害刺激を，標的音声との弁別が可能な話者の音声とする理由は，連続して標的音声と妨害刺激を聴いても互いに弁別ができなかった場合，後の記憶テストでも当然両刺激を弁別できず，標的音声を正しく同定できないことが予想される。そのような状況で得られた結果は，記憶された言語化情報が一致している程度による影響なのか，それとも元々弁別ができなかったことによる影響なのかがわからず，これらが交絡してしまう可能性があるからである。

本研究の目的

音声ラインナップを構成する音声刺激の選定方法は，従来の研究では，評定者によるおそらくは全体的な印象としての類似度の評定を基に行われていた（e.g., Olsson et al., 1998; Orchard & Yarmey,1995; McDougall et al., 2015; Vanags et al., 2005 ; Yarmey, 2001）。しかし，音声刺激間の類似性は，実験結果に影響する重要な変数となり得るため，特に言語化の有無を要因とする研究では，音声刺激の選定は慎重に行う必要がある。

そこで，本研究では，音声に関する幾らかの次元（例えば，「声の高さ」「声の太さ」）について言語化を行った類似度を指標として，「標的音声」と，その標的音声の言語化情報との類似度が高い「高類似音声」およびその類似度が低い「低類似音声」の選定を行う。そのために，各音声に対していくつかの形容詞対を用いた評定を行い，音声刺激間で類似している音声刺激（高類似音声），あるいは類似していない音声刺激（低類似音声）の候補を挙げるために多次元尺度構成法とクラスター分析を行う。同時に，その刺激の候補が十分に弁別可能なものであることを確認するために，弁別課題を用いる。弁別課題では，二つの音声刺激を連続して聴いてもらい，それら二つの話者が同じかあるいは異なるかという話者の同異判断を求める。なお，形容詞対には，木戸・粕谷（2001）が抽出した，通常の発話における声質に関連する表現語として一般性を持つ形容詞（例：「高い声–低い声」「落着きのない–落着きのある」）を用いる。

方法

参加者　聴覚に異常がない大学院生・大学生 16 名（男性 10 名・女性 6 名，19—23 歳）が，個別に実験に参加した。なお，本研究は，法政大学文学部心理学科・心理学専攻倫理委員会による研究倫理審査を受け，承認を得た（承認番号 19-0005）。

装置　音声刺激の録音は，Roland 社製のボイスレコーダーR-09HR（96kHz／24bit）で行った。WAV 形式に変換，音声刺激の音量と長さの調整は Praat（Boersma & Weenink, 2018）で行った。そのようにして作成した WAV 形式の音声刺激を，Apple 社製のノートパソコン MacBook Pro（44kHz／24bit）で再生し，ELECOM 社製のヘッドホン HS-HP13（20kHz）から出力した。なお，本研究は防音室で行った。

刺激　8 名の標準語話者の男性（23—26 歳）の音声を使用した。それらの音声の録音時には，自然に発話を心がけるように指示した。そのうち，1 分間のオレオレ詐欺を模したセリフ（Table 1 参照），14 種類の短い

セリフ（例：よろしくお願いします；Table 2 参照）および 1 分間のアリバイ証言を模したセリフ（Table 3 参照）を使用した。

手続き　まず，参加者には，本研究の目的と方法，参加の自由，同意の撤回ができること，プライバシーへの配慮等を予め文面と口頭で伝えた上で，本実験の参加に同意するかどうかの回答を求めた。同意した参加者には，評価段階として，1 分間のオレオレ詐欺を模したセリフ（Table 1 参照）の音声を提示している間に，評価用紙に回答してもらった。評価項目の内容は，木戸・粕谷（2001）が行った聴取評価を参考にした 6 つの声質表現語対であり，例えば，「高い声–低い声」「落着きのない–落着きのある」といったもの（Table 4 参照）で，それぞれ「非常に（3）」，「かなり（2）」，「やや（1）」，「普通（0）」のいずれか当てはまる数字 1 つに〇を付けてもらった（7 件法）。6 つの声質表現語対は 1 枚の用紙に印刷されており，参加者は話者 1 人ごとに 1 分間の音声が提示し終わるまでに 6 つの項目全てに回答するように求められた。このようにして，8 名の話者の音声全てに対して聴きながら評定してもらった。音声刺激の提示順序は，8 通りのラテン方格により参加者間でカウンターバランスした。

Table 1
1 分間のオレオレ詐欺を模したセリフの内容

もしもし。
うん，イチロウだけど。
ちょっと風邪引いちゃってさ。
声がおかしいけど，気にしないでね。
そういえば，スマホ壊れちゃったから，新しいのに変えたんだ。
番号教えるから，登録よろしくね。09012345678。
あとさ，ちょっと頼みがあるんだけど……いいかな？
実は，友達の連帯保証人になってたんだけど，その友達がいなくなっちゃってさ。
代わりにお金を返す羽目になったんだけど，まだ用意ができてなくて……。
今日中に，210 万払うことになってるんだけど，
ちょっと立て替えてくれないかな？
え，いいの？
じゃあ，用意ができたら，富士見銀行の普通口座，名義がタナカミノル，口座番号が 135790 に振り込んでね。
本当にありがとう。

その後，同異判断段階として，2 つの異なるセリフの短い音声（Table 2 参照）を連続で呈示し，それらの音声の話者が同一人物である場合は「はい」，そうでない場合は「いいえ」と，口頭で回答してもらい，実験者がそれを記録した。2 つの音声は，同一人物のものである場合と，そうでない場合があった。異なる話者の組み合わせが 28 通りあり，それぞれについてどちらを先に提示するかを入れ替えて 56 通りとなった。その割合を等しくするため，同じ話者を組み合わせて 56 試行とし，1 人の参加者につき 112 試行行った。例えば，話者 1 と話者 2 を呈示するときは，話者 1「あけましておめでとう」と話者 2「ただいま帰りました」となった。なお，ちょうど半分の 56 試行目で小休憩（約 1 分間）をとった。

Table 2

14 種類の短いセリフの内容

了解，承知しました
予約をキャンセルします
末長く，お幸せに
本当に，すみません
大変恐れ入ります
少々お待ちください
よろしくお願いします
ただいま帰りました
そうなんだ，知らなかった
さようなら，また会おう
こんにちは，はじめまして
お電話代わりました
お先に失礼します
あけましておめでとう

次に，言語化段階として，参加者には，8 名中 1 名の話者の 1 分間のアリバイ証言を模したセリフ（Table 3 参照）の音声を聴いてもらった後，その音声の特徴や印象を自由記述で回答してもらった。その際，「本日の実験の最初のセッションで，評定に使用した形容詞を「使わなくてはならない」あるいは「使ってはいけない」ということはありませんので，思った通りに自由に表現してください」と書面と口頭で教示し， 5 分間でできるだけ自由に記述してもらった。なお，言語化段階は評定の妥当性を確認する手段として実施されたが，十分に根拠となるデータが得られたため，本論文では結果について報告しない。

最後に，「差し支えなければ」と伝えた上で，年齢や性別を回答してもらい，謝礼のお菓子を手渡し，実験終了とした。なお，所要時間は約 20 分であった。

Table 3

1 分間のアリバイ証言を模したセリフの内容

その日のことはよく覚えているよ。
確か……目覚まし時計がなる前に，目が覚めちゃったんだよな。
二度寝して，会社に遅刻しちゃったらいけないと思って，すぐに出勤する準備を始めたんだよ。
いつも通り，歯を磨いて，顔を洗って，ニュース番組を見ながら，朝食を取ったな。
その日の朝食は，確か……パン，だったかな？
パンを食べたらすぐに，家を出たよ。
駅に向かう途中で，ランニングしている人を見かけたな。でも，いつも見かける，犬と散歩している人は見かけなかったな。その日は雨が降っていたから，そのせいかもしれない。
駅に着いたら，もう電車がホームに来ていたんだけど，サラリーマンと学生で満員になっていたから，その電車は見送って，次に来た電車に乗ったんだよな。
会社に着いたらすぐに，仕事に取り掛かったんだけど，その日は忙しくて，出勤から退社まで，ずっとパソコンの前だったな。だから，誰かと電話する時間なんてなかったよ。

Table 4
話者別の各声質表現語対の評定値の平均値

話者	声質表現語対					
	低い-高い	かすれた-澄んだ	落着きのない-落着きのある	弱々しい-迫力のある	細い-太い	張りのない-張りのある
1	4.63	5.19	4.75	3.88	4.00	4.63
2	2.31	2.63	4.75	2.88	4.75	2.94
3	3.19	3.38	4.06	2.31	3.44	2.44
4	4.81	5.19	3.88	4.19	3.94	5.13
5	2.00	3.88	5.94	2.38	4.25	2.94
6	5.31	2.94	3.38	2.63	3.00	3.19
7	5.19	4.19	3.38	2.94	3.19	4.25
8	3.94	3.94	4.75	4.44	5.00	3.75

注）最小値は 1，最大値は 7 をとる。最大値に近づくほど，その音声が下段の形容詞（例えば，「低い–高い」の場合「高い」を指す）により説明できることを表す（例えば，「低い–高い」の場合「非常に高い」ということになる）。

結果と考察

本研究では，ある音声刺激間の言語化情報が類似している，あるいは類似していないことを検討するだけでなく，実際に連続して聴いた場合に弁別可能な音声刺激を選定することを目的としていた。

まず，話者ごとに算出した声質表現語対の評定値の平均値を Table 4 に示す。それを用い，言語化情報の類似度が互いに高い，あるいは低い音声刺激の候補を挙げるために，多次元尺度構成法（ALSCAL）と階層クラスター分析（ward 法）を行った。その結果，まず，多次元尺度構成法により，話者 1 と話者 4，話者 6 と話者 7，話者 2 と話者 5 の各ペアが，同じ象限に収まったことで，形容詞の評定値の類似度，すなわち言語化情報の類似度がそれぞれ互いに高いことが示された（Figure 1 参照）。さらに，階層クラスター分析により，話者 1 と話者 4 が最初の段階でクラスターに構成される刺激の組み合わせになっており，話者 6 と話者 7 がその次の段階のクラスターに構成されることが明らかになった（Figure 2 参照）。これらの分析により，話者 1 と話者 4，話者 6 と話者 7 の各ペアを，言語化情報の類似度が互いに高い音声刺激の候補とした。そして，弁別が十分に可能な音声刺激のペアを選定するために，同異判断段階の弁別課題の正答率を算出したところ，話者 6 と話者 7 は 43.8%と低かったが，話者 1 と話者 4 は 81.3%と十分に高かった。そのため，話者 1 と話者 4 を言語化情報の類似度が互いに高い音声刺激とした。

次に，多次元尺度構成法の分析結果を示した Figure 1 で，話者 1 および話者 4 と異なる象限にあり，それらと最も距離がある，話者 3 を言語化情報の類似度が低い音声刺激とした。また，話者 1 と話者 4 のうち，話者 3 と弁別が十分に可能な方を「標的音声」とするために，同異判断段階の弁別課題の正答率を算出したところ，話者 1 と話者 3（78.1%）より，話者 4 と話者 3（87.5%）の方が弁別課題の正答率が高かった。従って，話者 4 を「標的音声」とし，その標的音声との言語化情報の類似度が高い話者 1 を「高類似音声」，逆にその類似度が低い話者 3 を「低類似音声」とすることとした。

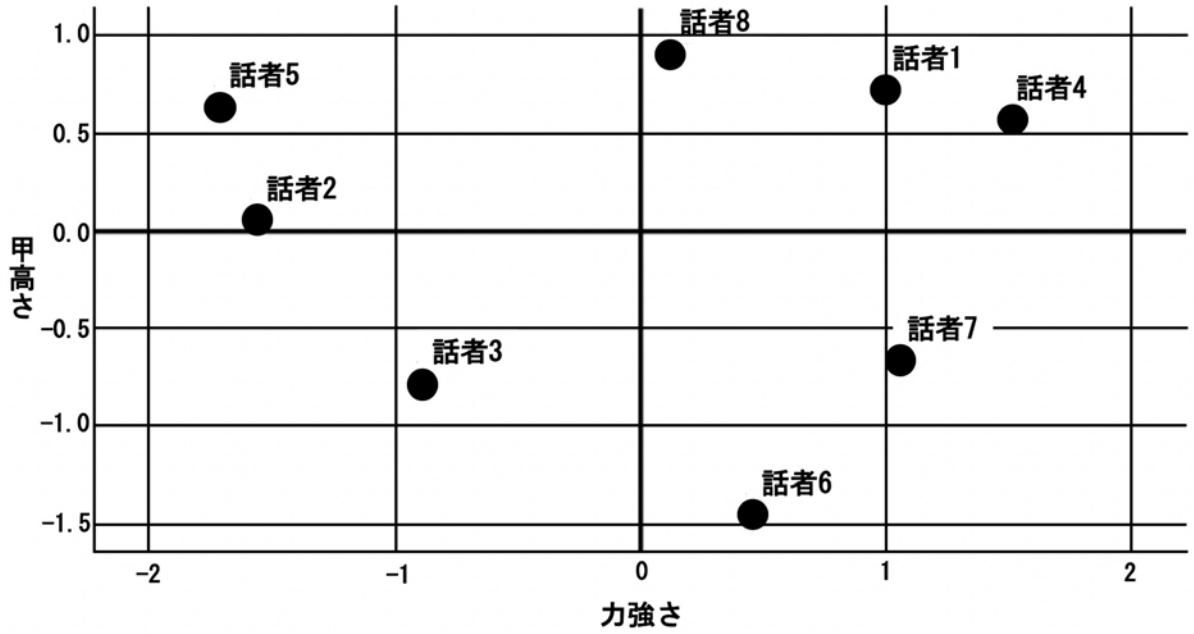

Figure 1. 多次元尺度構成法（ALSCAL）により得られた，声質表現語対の評定値に基づく話者の 2 次元布置。

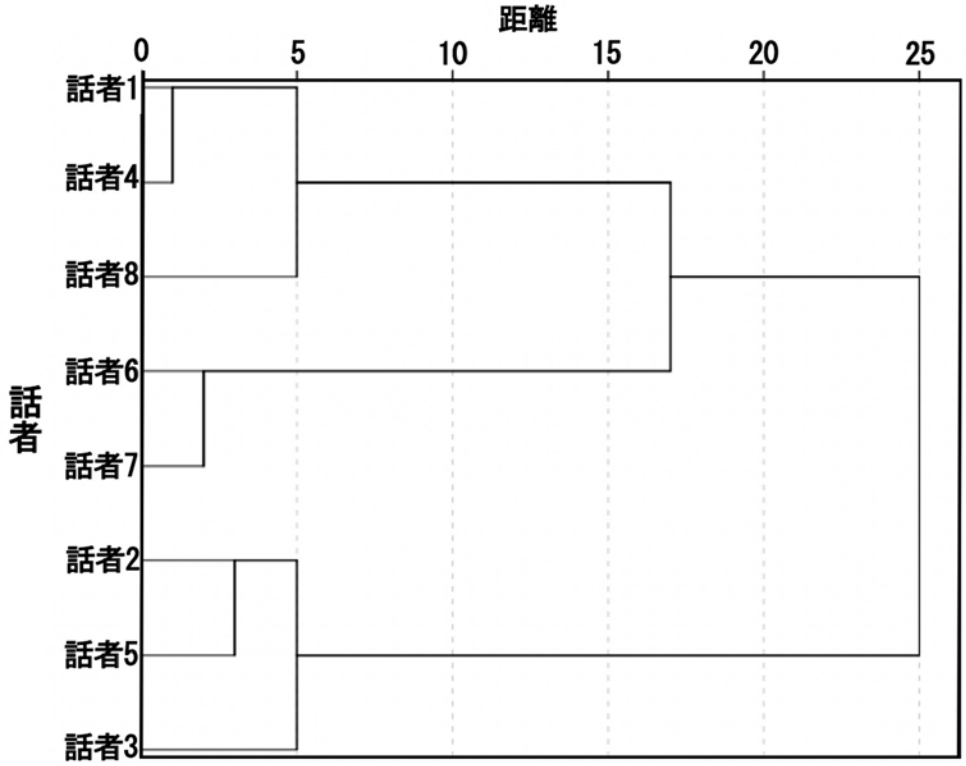

Figure 2. クラスター分析により得られた，声質表現語対の評定値に基づく話者のデンドログラム。

本研究は，標的音声の特徴や印象（例えば，声の高さや太さ）を言語的に符号化することで生み出された言語的な情報，すなわち「言語化情報」の類似性により妨害刺激を選定した。これは，Vanags et al.（2005）などの従来の研究で採用していた全体的な「聴いた印象」の類似性が低い妨害刺激が，必ずしも標的音声との「言語化情報」の類似度も低いとは限らないという疑念から提案した方法であった。実際，本研究で作成した音声刺激を用いた研究（井上，2020）により，今回作成した音声刺激が機能的に異なることが証明されている。

井上（2020）では，話者同定テスト時に「標的音声」，標的音声との言語化情報の類似度が高い「高類似音声」および，その類似度が低い「低類似音声」を聴くごとに，「（学習時に聴いた）標的音声と同じ人物だと思う程度」をそれぞれの音声刺激に対して個別に回答してもらった。その結果，まず，高類似音声については，統制群では，標的音声のほうが高類似音声よりも「標的音声と同じ人物だと思う程度」の評定値が有意に高かった一方で，言語化群では，標的音声と高類似音声に「標的音声と同じ人物だと思う程度」の評定値に有意な違いがみられなかった。これは，高類似音声は聴いた印象は標的音声とは類似していないが，言語化情報は標的音声と類似している音声刺激であることを示す。すなわち，高類似音声は標的音声の言語化情報との類似度が高い妨害刺激として，正しく選定できていたといえるだろう。

次に，低類似音声については，統制群と言語化群ともに，標的音声のほうが低類似音声よりも「標的音声と同じ人物だと思う程度」の評定値が高かった。これは，低類似音声は聴いた印象でも言語化情報でも，標的音声とは類似していない音声刺激であることを示す。すなわち，低類似音声も標的音声の言語化情報との類似度が低い妨害刺激として，正しく選定できていたといえるだろう。

また，以上のことから，必ずしも標的音声との「聴いた印象」の類似度が低い妨害刺激が，標的音声との「言語化情報」の類似度も低いとは限らないことが指摘できる。従って，今後，言語化の有無を要因とした研究を行う際には，本研究のような，音声に関する幾らかの次元において言語化情報の一致度に基づき，標的音声と妨害刺激を選定することが望ましいだろう。

引用文献

Boersma, P., & Weenink, D. (2018). Praat: Doing phonetics by computer [computer program]. Version 6.0.40, Retrieved from http://www.praat.org/（2018 年 5 月 11 日）

井上 晴菜（2020）．音声の言語的符号化が話者同定の成績に与える影響――音声ラインナップの類似性に着目して――日本心理学会第 84 回大会

木戸 博・粕谷 英樹（2001）．通常発話の声質に関連した日常表現語――聴取評価による抽出―― 日本音響学会誌，*57*，337–344.

北神 慎司（2000）．視覚情報の記銘における言語的符号化の影響 心理学研究，*71*，387–394.

McDougall, K., Nolan, F., & Hudson, T. (2015). Telephone transmission and earwitnesses: Performance on voice parades controlled for voice similarity. *Phonetica*, *72*, 257–272.

Olsson, N., Juslin, P., & Winman, A. (1998). Realism of confidence in earwitness versus eyewitness identification. *Journal of Experimental Psychology: Applied*, *4*, 101–118.

Orchard, T. L., & Yarmey, A. D. (1995). The effects of whispers, voice-sample duration, and voice distinctiveness on criminal speaker identification. *Applied Cognitive Psychology*, *9*, 249–260.

Perfect, T. J., Hunt, L. J., & Harris, C. M. (2002). Verbal overshadowing in voice recognition. *Applied Cognitive Psychology*, *16*, 973–980.

Schooler, J. W., & Engstler-Schooler, T. Y. (1990). Verbal overshadowing of visual memories: Some things are better left unsaid. *Cognitive psychology*, *22*, 36–71.

田中 利樹（2017）．声の記憶の言語的符号化が話者同定に及ぼす影響――ターゲットボイスの長さに注目して―― 法政大学人文科学研究科修士論文（未公刊）

Vanags, T., Carroll, M., & Perfect, T. J. (2005). Verbal overshadowing: A sound theory in voice recognition? *Applied Cognitive Psychology*, *19*, 1127–1144.

Yarmey, A. D. (2001). Earwitness descriptions and speaker identification. *International Journal of Speech Language and the Law*, *8*, 113–122.

Yarmey, A. D. (2007). The psychology of speaker identification and earwitness memory. In R. C. L. Lindsay, D. F. Ross, J. D. Read, & M. P. Toglia (Eds.), *The handbook of eyewitness psychology, Vol. II: Memory for people* (pp. 101–136). New York: Psychology Press.

Yarmey, A. D., Yarmey, A. L., Yarmey, M. J., & Parliament, L. (2001). Commonsense beliefs and the identification of familiar voices. *Applied Cognitive Psychology*, *15*, 283–299.