

声に関するステレオタイプの解明に向けて ——音声に基づく人物像の知覚の3次元モデル——

勅使河原 三保子

キーワード：音声，印象形成，ステレオタイプ，3次元モデル，対人認知，動物の発声

1. はじめに

我々は普段、人の話し声を聞くと、その言語内容を理解するだけでなく、声や話し方（声の出し方や発音の仕方）から話し手の性格、身体的特徴、感情、出身地などに関する印象を何気なく、しかし瞬時に形成する。たとえば我々はまだ会ったことがない相手からかかってきた電話で、相手の話す用件に耳を傾けながらも、その相手の年齢や性別に始まり容姿、性格、そしてその時の気分などについて瞬時に想像することができる。このような印象は異なる聞き手の間でも驚くほど一致する一方で、話し手の実際の特徴とは必ずしも一致するわけでない。たとえば上記の例で言えば、実際に電話の相手に会ってみたら、声を聞いて抱いた印象とまるで違っていたという経験のある人も少なくないだろう。このように我々が話し方を聞いて抱く印象と話者の実際の特徴とが必ずしも一致しない原因は、我々が話者の話し方を聞いて形成する印象が声に関するステレオタイプ、すなわちこういう属性（性別、年齢、出身地、性格等）の人はこういう話し方をするというイメージとの、必ずしも事実と一致するとは限らない関係に基づいているからである。したがって、我々がどのような音声的特徴を持つ話し方を聞いてどのような印象を抱くのかという音声に基づく印象形成のメカニズムを解明するには、声のステレオタイプと音声的特徴の対応を明らかにする必要がある。

話し手の声や話し方が聞き手に与える印象に関する研究は、音声という研究対象が学際的であるため、様々な分野で行われている。音声・音響の分野でこの種の研究がさかんになったのは比較的最近かもしれないが、古くは心理学・行動科学の分野に始まり（先行研究の概観については Teshigawara [2003, 1.2.2 節] 参照）、その後サービス業、広告業、マーケティング (Ambady, Krabbenhoft & Hogan 2006; Chattopadhyay et al. 2003 他)、政治 (Klofstad, Anderson & Peters 2012 他)、医療 (Ambady et al. 2002 他) などの分野でも関心が持たれるようになった。また、扱うテーマも、サービス業や政治等でふさわしい声、どんな声・話し方の医師が医療ミスを訴えられる可能性が高いか、容貌・身体的特徴、性格特徴との関わり（声を聞いて容貌が魅力的な人や、身長・体重、年齢、性格が当てられるか）、など多岐にわたる。本論文では音声が聞き手に与える印象に関する先行研究を概観し、音声に基づくステレオタイプに関してより体系的な理解を深めるために仮説を立て、検証方法を提案する。

2. 音声が聞き手に与える印象

勅使河原 (2015) では、音声に基づく話者の属性の推定について、以下のようにならべている。まず、聞き手が話し声を聞いて話者の正しい年齢や性別、体格、性格や感情などを推定できるのか調査した研究によると、年齢や性別は声からある程度正確に推定できることがわかっている（たとえば年齢については Ptacek & Sander [1966]、性別については Perry, Ohde & Ashmead [2001] 参照）。感情については、話者がその感情を感じている最中の音声を入手する困難さから、感情を演じた「ふり」の音声をういて行う研究が多いが、その中でも「本物」の感情を表した音声にこだわった Erickson et al. (2006) の研究では、本当に悲しい体験を経た話者がその体験を回想して悲しみながら発話した音声（本物の悲しみの音声）と、後からその音声の抑揚を模倣した音声を収録し、本物の悲しみの音声と模倣の悲しみの音声を被験者に聞かせたところ、模倣の悲しみの方がより話者が悲しんでいる印象を与えたと報告している。このことから、我々

は話し声から話者の感情をある程度は正しく推定できるものの、我々の判断は声に関するステレオタイプにも影響を受けると考えられる(勅使河原 2015)。

一方、性格については、1930～40年代に聞き手が話者の性格を話し声からのくらい正確に推定できるかという研究が、その後、話者の性格検査の結果と声の様々な音響的特徴を関連付ける研究が行われたが、聞き手が話者の性格を当てたり、話者の性格と音響的特徴を関連付けたりするのは難しいことが判明した(勅使河原 2015)。代わりに話者の本当の性格とは関係なく、聞き手の間で性格に関する印象が一致する、声のステレオタイプに基づくと考えられる現象が繰り返し指摘された(Brown & Bradshaw [1985] 参照)。1980年代以降は話者の性格自体に言及せず、素人の聞き手に声の音声的な特徴と話者の性格に関する印象を尋ね、二者の相関を調べる研究(attribution study もしくは inference study) が一般的になっている(Brown & Bradshaw *ibid*)。すなわち、我々が話し声聞いて話者の性格を推定するのは一般的に不可能であるものの、聞き手が話者の性格について他の聞き手とも共通した印象を抱くのは、声に関する共通したステレオタイプを持っているからであると説明できる(勅使河原 2015)。

しかし、これらの研究の多くは、音声の専門家が行ったもの(内田 [2009, 2011] 等; 籠宮他 [2007a, 2007b] 等参照)を除き、たとえば評定項目間の相関についてといった声から受ける印象のみについては扱っても、印象をもたらす音声的な特徴についてほとんど議論していない。また、音声的な特徴に言及していても原始的で大雑把な特徴しか考慮しないものが多い(たとえば Hecht & LaFrance [1995], Yarmey [1993])。しかしながら、音声のどのような特徴が特定の印象に寄与するのか、聴覚に基づく印象形成のメカニズムを解明するためには、音声的特徴との関連を詳しく調査することが不可欠である。

一方で、内田(2009, 2011)は性格特性5因子論に基づいた Big Five と呼ばれる extraversion (外向性)、neuroticism (情緒不安定性)、openness (経験への開放性)、conscientiousness (勤勉性)、agreeableness (協調性)の五つの性格特性を評価次元として用い、イントネーションパターン、発話速度、フォルマント周波数などの音声的な特徴との相関を調べている。これらの緻密な実験計画・分析方法は十

分評価されるべきである。しかし一方で、初対面（あるいは電話などで初めて声を聞く場面）で音声を頼りに本能的に相手の人物像を形成するような場合、我々は果たして本当に 5 因子で代表されるような評価次元を用いるのだろうか。そうではなく、人が初対面で他人に対して印象を抱く対人認知において普遍的であるとされる 2 次元、Warmth（温かさ）と Competence（能力）のようにもっと生得的、直観的で基本的な次元が音声に基づく印象形成にも介在しているのではないか。（たとえば Fernández Gallardo & Weiss [2017] も音声に基づく印象形成における Big Five を基にした評価次元の活用について同様の疑問を抱いている。）このような考えの下、Teshigawara (2011) では、対人認知の Fiske, Cuddy & Glick (2007)、対人認知でも重要な役割を果たす顔の認知に関する研究 (Oosterhof & Todorov 2008)、さらに動物の発声に関する研究 (Morton 1994) を援用し、聴覚に基づく印象形成において重要な 2 次元とその音声の特徴を推測した。しかしながら、Teshigawara (2011) では、Teshigawara (2003)、Teshigawara et al. (2009) のデータを 2 次元に照らして再考察したのみで、これら 2 次元が新しいデータを基に検証されたわけではない。一方で、音声にはこれら 2 次元で扱われない、韻律的特徴を含む動的な音声的特徴も存在し、音声に基づく印象形成において重要な役割を果たしていることが考えられるため、2 次元モデルでは不十分である。そこで、本論文ではまず Teshigawara (2011) を援用し、さらに Osgood (1964) の三つ目の次元である Activity (活動性) を加え、音声に基づく印象形成の 3 次元モデルを提唱する。（なお、「活動性」は籠宮他 [2007a, 2007b] でも次元の一つとして言及されている。）次節では 3 次元モデルの詳細と、それに基づき、音声に基づく印象形成についての三つの仮説を立て、検証方法を検討する。

3. 音声に基づく印象形成のモデルと仮説

3.1 音声に基づく印象形成のモデル

3.1.1 2次元モデル

ここではまず、Teshigawara (2011) で提唱された音声に基づく印象形成の

2次元モデルを説明する。Fiske et al. (2007) は既存の対人認知に関する研究をまとめ、人が無意識に他人の行動を解釈し印象を抱く時の基本的な次元として warmth と competence を挙げている。これは他人と出会って即座に行う、① warmth (敵か味方か、すなわち持ち合わせる意図が善か悪か)、そして② competence (その意図を実行できる能力があるか否か) の二つの次元に関する判断が、対人認知において普遍的であることを示す。また、実際の判断においては warmth の判断が competence に先行し、さらに warmth の方が感情・行動的な反応においてより比重が大きいとしている。Fiske et al. によると、warmth の次元が valence、すなわち印象の好悪に相当し、competence の次元は、その印象がどれくらい極端であるか (たとえばどのくらい好ましいか) に相当する。これらの2次元は、長い研究の歴史の中で繰り返し現れ、様々な名称で呼ばれてきたが、流れを異にする小規模集団でのやりとりに基づく印象形成に関する研究でも、これらを含む3次元が出現している¹。なお、Teshigawara (2011) では、音声に基づく印象形成と共に対人認知の下位分野を形成する顔の認知の研究手法に関心を持ち、CGを用いた顔の研究である Oosterhof & Todorov (2008) で出現した次元 (valence と dominance) とその解釈についても言及している。これらの2次元は言葉こそ異なるが、より一般的な対人認知の分野で抽出された2次元と実質的に同じであると考えられる。

3.1.2 予測される音声的特徴

Teshigawara (2011) では、これらの2次元を対人認知の下位分野である顔の認知における知見を介して、鳥類・哺乳類の発声における2次元と比較し、これ

1 Fiske et al. (2007, p. 78) は Bales, R. F., ed. (1999) *Social interaction systems: Theory and measurement*, Transaction Press を引用し、3次元目がやりとりの量であったことに言及し、生のやりとりにおける印象評定では鮮明に浮き上がるが、記憶として保持された印象ではさほど際立たないだろうとしている。これは音声に基づく印象形成に文脈を置き換えると、音声自体が動的な性質を持つため、印象評定に用いられる比較的短い音声であっても、より鮮明に出現することが予測され、本論文で3次元目として Activity を提唱する動機づけともなる。

ら 2 次元の音声的特徴を予測している。Oosterhof & Todorov (2008) では、CG 画像を用い、valence と dominance の 2 次元の SD (標準偏差) を基に変化させた顔を刺激として用いた。すると valence の得点が高い顔に対しては happy、中間の顔に対しては neutral、低い顔に対しては angry の評定が統計的に有意に付された。刺激に用いた顔は全て感情的には中立で、各々の次元の評定に関係のある特徴を誇張したものであったことから、著者らは被験者が感情表現の知覚を過剰一般化して評定したと解釈した。一方、感情表現の知覚の過剰一般化では説明できなかった dominance の評定について、顔の男性らしさや成熟度が身体的な強さ、ひいては意図の実行能力の知覚につながるのではないかと考え、被験者に dominance の SD を基に変化させた顔の男性らしさと成熟度を評定させたところ、統計的に有意な相関が得られた。つまり、valence の評定は快・不快の感情表現との類似性に影響を受けるが、dominance の評定は顔の男性らしさ・成熟度の印象に影響を受けると結論付けている。この結論に基づくと、同様に対人認知の下位分野である音声に基づく印象形成においても warmth/valence の知覚は快・不快の感情表現との類似性に影響を受け、一方 competence/dominance の知覚は声の男性らしさや成熟度の印象に影響を受けると仮定することが可能となる。

次に Teshigawara (2011) では、音声学者である J. J. Ohala が Morton (1977) が鳥類・哺乳類の発声について文献調査を行いまとめた研究 (Ohala 1984, 1994) に基づき、人間および動物の音声コミュニケーションに関して提唱した Frequency Code² に着想を得て、音声に基づく印象形成の 2 次元の音声的特徴

2 Frequency Code とは、人間および動物が、自分が小さくて相手に対して脅威を与えない存在であることを示す時に高い基本周波数 (F0) やフォルマント周波数を用い、逆に自分が大きくて相手に対して攻撃的、支配的であることを示す時に低い F0 やフォルマント周波数を用いるという対応関係を表したもので (Ohala 1984, 1994)、Morton の MS 規則の第 2 次元と関連する (Morton 1994)。この前提にあるのは、F0 は発声する個体が大きければ大きいほど低いのが一般的であるため、F0 は体の大きさに反比例すると考えられるからである。ただし、実際には F0 と体の大きさとの相関はそれほど良くなく、代わりにフォルマント周波数こそが体の大きさとより良い相関を示す報告もある (Fitch 1997)。Frequency Code はその後音声学界で大きな影響力を与え続けている。

を Morton (1977, 1994) に基づいて予測している。

Morton (1977) は鳥類・哺乳類の①相手が近くにおいて、攻撃的で (aggressive) 相手から距離を取りたい時と、②争いを静め友好的な関係を保ち (appeasing, friendly)、相手との距離を縮めたい時、の二つの場面における発声の種類傾向をまとめ、これら二つの場面 (意図) での動物の発声 (信号の構造) を説明する規則として Motivation-Structural Rules (以下、MS 規則) を提唱した。MS 規則によると、動物は攻撃的な場面ではピッチが低くざらざらとした耳障り (harsh) な発声をし、友好的な場面ではピッチが高くより周期的 (tonal) な発声をする と一般化できる。Morton (1994) では、音声の特徴がさらに整理され、第 1 次元として音の質 (ざらざらした非周期的な発声から周期的な発声まで様々に変化)、第 2 次元として F0 あるいはスペクトルのエネルギー分布としている。(さらに第 3、4 次元として各々速さと音量を挙げているが、文献ではそれ以上の言及はない。) MS 規則の第 1 次元である「質」について Morton 自身はあまり第 2 次元と切り離しては言及していないが、Fichtel, Hammerschmidt & Jürgens (2001) はリスザルを対象に快・不快の時に発した鳴き声の音響分析を行い、不快の時の発声の方がより非周期性が強いことを報告している。

Teshigawara (2011) では、これらの動物の発声に関する先行研究に基づき、動物の発声に関する MS 規則を対人認知における二つの次元と比較し、動物の発声においても第 1 次元 (声の質) が意図の善し悪し (warmth/valence) を表し、顔の認知の場合は感情表現 (happy - neutral - angry) と相関があったのが、動物の発声においても快・不快の感情と関連があると仮定している。次に第 2 次元 (周波数情報) を基に提案された Ohala の Frequency Code が体の大きさも表すことから (脚注 2 参照)、顔の認知での成熟度にもつながる身体的な強さとも関係づけられ、意図の実行可能性 (competence/dominance) と置き換えられると考える。よって、音声に基づく印象形成の 2 次元を対応する音声的特徴と共に表すと、以下ようになる。

第1次元：意図の善し悪し（音声的特徴：快・不快の感情を区別する声質の非周期性）

第2次元：意図の実行可能性（音声的特徴：体の大きさと負の相関のあるF0とフォルマント周波数）

Teshigawara (2011) では、Teshigawara (2003)、Teshigawara et al. (2009) の結果を再考察し、印象評定と音声的特徴との間に現れた相関関係と2次元との比較を行っている。それによると、①好ましい特徴を表す評定項目（「ハンサム・美人だ」、「誠実な」等）の間に正の相関が見られ、②「(体が) 大きい」と「強い」の間に正の相関が、これらと「(声が) 高い」の間に負の相関が見られ、①・②が上記の2次元に対応することが示唆されている。また、第1次元に相当する好ましい特徴を表す評定項目は著者の受聴分析によって刺激音に付与された「咽頭部の狭め」(harsh voice や creaky voice を発声する時に起こる) の程度と負の相関があったが、咽頭部の狭めは Scherer (1986) の不快感情を表す音声特徴の予測とも一致し、MS 規則に基づき予測された、聴覚に基づく印象形成の第1次元（感情の快・不快を区別する声質）とも一致するとしている。一方、第2次元に相当する「(体が) 大きい」、「強い」等の項目群はF0平均とはあまり相関が高くなかったものの、母音「あ」のF2平均と負の相関があったことに基づき、第2次元（体の大きさと負の相関のあるフォルマント周波数）と一致するとしている。さらに、Teshigawara et al. (2009) ではヘブライ語母語話者を被験者として同一刺激音を聞かせて行った実験結果も報告されており、上記①・②の傾向については日本語母語話者の結果と概ね同じであったことから、これら聴覚に基づく印象形成の2次元の言語・文化の違いを超えた普遍性が示唆されるとしている。しかしながら、Teshigawara (2011) では自身の携わった過去のデータの再考察を行ったのみで不十分であるため、新しい独自のデータを基にこれらの次元を検証する必要があるのは自明のことである。

翻って対人認知や顔の認知の研究に基づく予測を立てて研究を行ったという点で本論文と共通点を持ちながらも音声学ではなく心理学の専門家が行った最

近の研究 (McAleer, Todorov & Belin 2014) では、64 人の話者が電話の会話を想定し、“hello”と発話した 0.5 秒未満の音声を 320 人の被験者に評定させた実験から、対人認知や顔の認知の研究と比較可能な Valence と Dominance の 2 次元を抽出している。McAleer et al. はこれらを軸とした 2 次元空間を social voice space と名付けている。測定した様々な音響特徴量のうち F0 とフォルマント周波数、すなわち本モデルの第 2 次元の音声的特徴が両次元の分散を部分的に説明できることが判明したものの、理論的な説明は見られない。このように McAleer et al. の研究結果は、本論文のモデルの音声的特徴の一部が 2 次元に関連があることを示唆するものの、音声に基づく印象形成のメカニズムが十分に解明されているとは言い難く、本論文で二つの次元を音声的特徴の予測と共に示すことに意義があると言えよう。

しかしながら、これら 2 次元で扱われる特徴は声質の非周期性と F0・フォルマント周波数のみで、すでに指摘したように音声に基づく印象形成において重要な役割を果たしていると思定される韻律的特徴を含む動的な音声的特徴が扱われないため、2 次元モデルでは不完全である。したがって、本論文ではこれらに Activity (活動性) というもう一つの次元を加えた 3 次元モデルを以下に提唱する。

3.1.3 3次元モデル

C. E. Osgood と同僚らは 1950 年代より形容詞の対を用いて被験者に様々な評価対象を評定させたデータを分析してきた中で、以下の三つの因子が繰り返し現れるのを確認した (Osgood 1964) : ①良い—悪い、快—不快、肯定的—否定的などに代表される評価の因子 (Evaluative Factor) ; ②強い—弱い、重い—軽い、硬い—軟らかいなどに代表される力量の因子 (Potency Factor) ; そして③速い—遅い、能動的な—受動的な、興奮しやすい—冷静ななどに代表される活動性の因子 (Activity Factor)。これらのうち、初めの二つは Teshigawara (2011) が提唱する第 1 次元「意図の善し悪し」(warmth/valence) および第 2 次元「意図の実行可能性」(competence/dominance) に概ね相当すると考えられる。

また、音声の動的な特徴を扱うためには Osgood の 3 番目の因子である「活動性」(activity) をモデルに取り入れるのが理に適っている。この「活動性」という因子は講演音声の評定したデータと大局的な音声的特徴の相関を研究した籠宮他 (2007a, 2007b) でも出現する因子であるが、籠宮他はポーズ比が「活動性」と相関があったと報告している。音声の動的特徴を捉えられる音声的特徴にはその他ピッチ曲線、発話速度等も考えられる。よって、Teshigawara (2011) の音声に基づく印象形成の 2 次元モデルは以下のように改訂できる。

第 1 次元：意図の善し悪し (音声的特徴：快・不快の感情を区別する声質の非周期性)

第 2 次元：意図の実行可能性 (音声的特徴：体の大きさと負の相関のある F0 とフォルマント周波数)

第 3 次元：活動性 (音声的特徴：ピッチ曲線、発話速度、ポーズ比等)

そして、これらを図示すると図 1 のようになる。

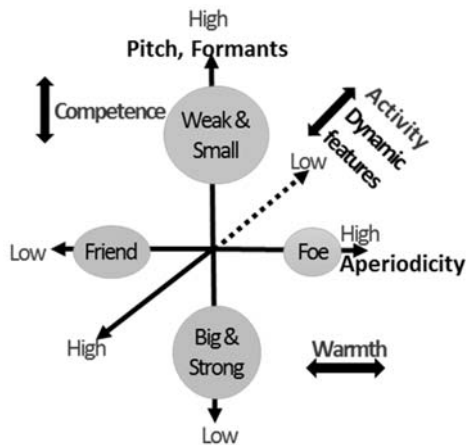


図 1：音声に基づく印象形成の 3 次元モデル

3.2 仮説と検証方法

前節では本論文が依拠する音声に基づく印象形成の3次元モデルを導入したが、これらの次元のうち第1、2次元は対人認知の異文化間比較の研究も交えた先行研究に基づいて提唱されており (Fiske et al. 2007)、第3次元の活動性もまた異文化間の比較を交えた研究の結果によるものである (Osgood 1964)。また、予測される音声的特徴には種の違いを超えて動物の発声に観察された特徴に基づくものが含まれるが (第1・2次元)、そのうち第2次元は Ohala (1984, 1994) の Frequency Code によって言語音声においても検証されている。よって、複数の言語・文化において音声を聞かせて印象評定をさせる実験を行ったら、言語・文化の違いを超えてこれらの3次元が出現すると予測するのは道理にかなっていると考えられる³。この共通性は3次元モデルが依拠する対人認知における言語・文化の違いを超えた普遍性や人間と動物の発声との共通性から来るものである可能性もちろんあるが、一方でこれだけマスメディアが発達した現代においては、たとえばハリウッド映画のような全世界に対して影響力のあるコンテンツが発信するステレオタイプが言語・文化の違いを超えて共有された結果である可能性を即座に否定することはできない。したがって、これらの3次元がマスメディアによる声に関するステレオタイプの流布による結果ではなく、言語・文化の違いを超えて人間に生まれつき備わったものによる可能性が高いことを示すためには、単に異なる言語・文化の間で比較するだけでは不十分で、マスメディアへの接触が限定的である民族も調査の対象にする必要があると考える。ここまでの議論をまとめると、一つ目の仮説は「音声に基づくステレオタイプには言語・文化の違いを超えた普遍性が存在し、それは三つ

3 その他、対人認知モデルとは無関係で声に関するステレオタイプの異文化間比較を行った van Bezooijen (1988) では、異文化間で声質に関する音声的特徴と印象評定項目の相関の向きが同じであったため、本モデルの第1次元を支持すると考えられる。一方で、Shayan, Ozturk & Sicoli (2011) によると、ペルシャ語、トルコ語、サボテク語においては、声の高低を表す語彙が各々強弱に対応する英語とは異なり、各々「細い」、「太い」に対応する。本研究の刺激音で第2次元との対応が予測される音声的特徴 (F0、フォルマント周波数) の制御にはより注意を払う必要があるであろう。

の次元で説明できるとすることができ、聴取実験はマスメディアへの接触が限定的である民族も含め、言語・文化の異なる被験者を対象に行うべきである。

なお、本論文で提唱する3次元は音声に基づく印象形成に関する先行研究だけでなく、対人認知や動物の発声における研究に基づいているが、実際に聴取実験を行う際には、各次元を構成すると想定される複数の形容詞句を評定項目として用いることとなる。そのためには、仮に理論的に妥当だったとしても初めから恣意的な形容詞句を用いて被験者に音声の評定をさせるのでは、本研究が目指す我々人間にとって直観的で本質的な次元の抽出を阻害する可能性がある。そこで、本研究では Oosterhof & Todorov (2008) の顔の認知の研究手法に倣って、まず予備的に行う聴取実験で音声聞かせ、瞬時に想起される性格などの特徴を自由に記述させ、そこから得られる語句から合理的な方法で音声に基づく印象形成において有効な語句を導き出し、それらを用いて改めて評定させた結果を主成分分析などの統計的処理にかけると望ましいと考える⁴。

一方で、声に関するステレオタイプには言語・文化による特殊性が存在する可能性が十分にある。その特殊性をもたらす要因としては、まず各言語の音韻体系（音素の数と種類、音節構造）、アクセント体系（強勢アクセント、高低アクセント、声調）、リズム体系（強勢拍、音節拍、モーラ拍）が考えられる。たとえば、世界の諸言語でもあまり頻度が高くない音素（口蓋垂音、入破音等の子音や円唇前舌母音、鼻母音等）の有無や、許容される音節構造の違い（開音節 (CV) が中心の日本語のような言語に対して、連続できる子音数が多い言語（たとえば英語では最大で CCCVCCCC））は、他言語を母語とする聞き手に対して異なる印象を与えることは容易に想像できる。そこで、研究の対象として

4 Oosterhof & Todorov (2008) は顔の判断において基本となっている次元を特定するため、被験者に感情的に中立な顔から瞬時に想起される性格などの特徴を自由に記述させ、その自由記述で得られた特徴をまず 14 の次元（項目）に統合した。次に、その 14 項目に、対人認知に関する研究で重要である **dominance** を加えた 15 の項目を用いて同じ顔を評定させ、結果を主成分分析にかけた。そして、第 1 成分は、好ましい特徴が正の、好ましくない特徴が負の負荷を示したため、**valence** とし、第 2 成分は **dominance** に加え、**aggressiveness** や **confidence** なども負荷を示したので、**dominance** と名付けた。

選択する言語には系統的に異なり、できるだけ音韻体系やアクセント体系、リズム体系にばらつきが出るようなものを選ぶのが望ましいと考える⁵。

さらにここで大胆な仮説を立てると、同じ広さの母音空間を用いて音声コミュニケーションがなされるならば、母音音素が少ない言語の方が多言語に比べ、音素一つ辺りが実現される時の音声の自由度が広く、それが言語情報以外の意図や態度、さらに演出したい人物像の表現にも用いることが可能になるとは考えられないだろうか。たとえば日本語の「ア」/a/ は話者や文脈等によっては IPA で表記すると [a, æ, ɐ, ɑ, ʌ, ə] 等の音声で実現される可能性があるが、その母音空間が四つの音素 (/æ, ɑ, ʌ, ə/) で占められる英語とでは、音声実現の自由度に違いがあると考えてよいだろう。すると、日本語のような母音が多くない言語においては、その自由度が音声コミュニケーションにおいて印象（たとえば、自分は良い人間で脅威ではないという）を作り出すのに貢献している可能性はないだろうか。

また、同様に子音においても、たとえば日本語のラ行子音（流音）は 1 種類しかなく、その多くは弾き音 [r] で実現されるものの、4 人の日本語母語話者の日本語発話における 1500 以上のラ行子音の発音を丹念に受聴分析により記述した Magnuson (2009) によると、英語の /l, r/ (すなわち側面接近音と後部歯茎接近音) も含めた様々な発音が実現された。さらに、たとえばふるえ音 [r] (いわゆる巻き舌) は江戸言葉のべらんめえ口調⁶ で使われると言われ、筆者もテレビドラマ等の悪役男性等が使うのを観察したことがある他、一貫してラ行子

5 JSPS 科研費 18K00547 の助成を受けて行われる本研究では、研究対象として日本語、英語、そしてマスメディアへの接触が限定的なアフリカのフラニ族が話すフラニ語を第一段階で扱う。第二段階ではさらに口蓋垂音や咽頭音があることで特徴的なアラビア語、子音音素が多く鼻母音もあるヒンディー語、母音音素が多く声調言語であるベトナム語、子音母音共に平均的な音素数で声調言語である中国標準語、最後に円唇前舌母音を有し高低アクセントのあるスウェーデン語を扱う計画である。調査予定の言語が話される国・地域は広範囲にわたり、文化的にも多様である。

6 松村明（編）（2006）『大辞林第三版』で、べらんめえ口調は「早口で威勢がよく、巻き舌で荒っぽい言葉つき。江戸の下町で、主に職人などの間で発達したものと解説されている。

音を [l] で歌う歌手も観察される⁷。このうち特にふるえ音 [r] は一定の人物像（江戸っ子や悪役）との結びつきもあり、日本語母語話者にとっては特定の印象を与える可能性があるが、他言語を母語とする聞き手には恐らくそのような結びつきはなく、日本語母語話者に対してとは異なる印象を与えるだろう。このような音素レベルのバリエーションは3次元モデルの音声的特徴には含まれないが、実際に聴取実験を行い、もたらされる印象の差が体系的に観察されるならば、言語（この場合日本語）の特殊性としてモデルに組み込まれるべきであろう。ここまでをまとめると、二つ目の仮説として、「音声に基づくステレオタイプの言語・文化による特殊性にはその言語の音韻体系、アクセント体系、リズム体系が寄与している」とすることができる。

次に超分節的特徴の観点から見てみると、特定のイントネーションや声質が異なる言語、方言や文化によって好んで、あるいは慣習的によく用いられ、それが他言語・他方言を母語とする聞き手に対して独特の印象を与える一方で、その同じ特徴が聞き手の母語で別の文脈等で用いられる場合は、両言語で異なる印象を与えることにつながる可能性もあるだろう。たとえば、日本語においては「りきみ」の音声で恐縮した態度、すなわち最も丁寧度の高い発話で見られるが、それがフランス語母語話者には逆に無礼、権威的、いらだちとして知覚された例もある（Shochi, Aubergé & Rilliard 2005）。りきみ音声の場合は本論文で提唱する第1次元の音声的特徴が声質の非周期性であるため、言語間で比較した時、他次元に対する第1次元の配置の違いや、他次元との集約等によって現れる可能性も予測できる。翻って、ここで声質を発声様式に限定せず、舌や口唇などの音声器官を一定の構えに保つことでもたらされる質全般として扱い（Abercrombie [1967]、Laver [1994] の各々 voice quality, setting の定義参照）、硬口蓋化（palatalization）について見てみたい。硬口蓋化は多くの言語で共通し

7 神山（2008, p.35）では歌手の桑田佳祐氏について言及されているが、筆者の観察では太川陽介氏が「Lui-Lui」という楽曲にて一貫してラ行子音を [l] で発音するのが確認されている。しかし、一般の発話でラ行子音が一貫して [l] で発音される場合の聴覚印象については筆者の知る限り学術的には議論されていない。

て観察される音韻的プロセスとしてだけでなく、音韻的な条件付けがない環境における変化で「小ささ」を表す音表象 (sound symbolism) としても知られている (Ohala 1994)。発話全体にわたって硬口蓋化した声質で話す (すなわちフォルマント周波数が高めである) 話者が、体格が小さく、非力な印象を与えることは3次元モデルの第2次元で予測され、これは音表象とも関連すると考えることができる。よって、三つ目の仮説として「音声に基づくステレオタイプの中には音表象と関連するものがあるかもしれない」とまとめる。

4. おわりに

本論文では対人認知、顔の認知、動物の発声の知見に基づいて Teshigawara (2011) において提唱された、音声に基づく印象形成の2次元モデルを出発点とし、3次元モデルを提唱した。さらに、そのモデルに基づいて仮説を立て、検証方法についても論じた。本研究では今後、計画を実行して系統も音韻の特徴等も異なる複数言語を対象に聴取実験を行い、モデルと仮説の妥当性を検証していく。

参考文献

- Abercrombie, D. (1967). *Elements of general phonetics*. Edinburgh: Edinburgh University Press.
- Ambady, N., Krabbenhoft, M. A., & Hogan, D. (2006). The 30-sec sale: Using thin-slice judgments to evaluate sales effectiveness. *Journal of Consumer Psychology*, *16*, 1, 4-13.
- Ambady, N., LaPlante, D., Nguyen, T., Rosenthal, R., Chaumeton, N., & Levinson, W. (2002). Surgeons' tone of voice: A clue to malpractice history. *Surgery*, *132*, 1, 5-9.
- Brown, B. L., & Bradshaw, J. M. (1985). Towards a social psychology of voice

- variations. In H. Giles & R. N. St. Clair (eds.), *Recent advances in language, communication, and social psychology* (pp. 144-181). London: Lawrence Erlbaum.
- Chattopadhyay, A., Dahl, D. W., Ritchie, R. J. B., & Shahin, K. N. (2003). Hearing voices: The impact of announcer speech characteristics on consumer response to broadcast advertising. *Journal of Consumer Psychology, 13*, 3, 198-204.
- Erickson, D., Yoshida, K., Menezes, C., Fujino, A., Mochida, T., & Shibuya, Y. (2006). Exploratory study of some acoustic and articulatory characteristics of sad speech. *Phonetica, 63*, 1-25.
- Fernández Gallardo, L., & Weiss, B. (2017). Towards speaker characterization: Identifying and predicting dimensions of person attribution. *Proceedings of Interspeech 2017*, 904-908.
- Fichtel, C., Hammerschmidt, K., Jürgens, U. (2001). On the vocal expression of emotion: A multi-parametric analysis of different states of aversion in the squirrel monkey. *Behaviour, 138*, 97-116.
- Fiske, S. T., Cuddy, A. J. C., & Glick, P. (2007). Universal dimensions of social cognition: Warmth and competence. *Trends in Cognitive Sciences, 11*, 2, 77-83.
- Fitch, W. T. (1997). Vocal tract length and formant frequency dispersion correlate with body size in rhesus macaques. *The Journal of the Acoustical Society of America, 102*, 2, 1213-1222.
- Hecht, M. A., LaFrance, M. (1995). How (fast) can I help you? Tone of voice and telephone operator efficiency in interactions. *Journal of Applied Social Psychology, 25*, 2086-2098.
- Klofstad, C. A., Anderson, R. C., & Peters, S. (2012). Sounds like a winner: Voice pitch influences perception of leadership capacity in both men and women. *Proceedings of the Royal Society B, 279*, 2698-2704.
- Laver, J. (1994). *Principles of phonetics*. Cambridge, UK: Cambridge University Press.
- Magnuson, T. J. (2009). *What /r/ sounds like in Kansai Japanese: A phonetic investigation of liquid variation in unscripted discourse*. M.A. thesis, University of

- Victoria, Canada. [<https://dspace.library.uvic.ca/handle/1828/1367>] (2019年6月12日取得)
- McAlear, P., Todorov, A., & Belin, P. (2014). How do you say ‘hello’? Personality impressions from brief novel voices. *PLoS ONE* 9 (3): e90779. doi:10.1371/journal.pone.0090779
- Morton, E. S. (1977). On the occurrence and significance of Motivation-Structural Rules in some bird and mammal sounds. *The American Naturalist*, 111, 981, 855–869.
- Morton, E. S. (1994). Sound symbolism and its role in non-human vertebrate communication. In: Hinton, L., Nichols, J., Ohala, J.J. (eds.), *Sound symbolism* (pp. 348–365). Cambridge: Cambridge University Press.
- Ohala, J. J. (1984). An ethological perspective on common cross-language utilization of F_0 of voice. *Phonetica*, 41, 1–16.
- Ohala, J. J. (1994). The frequency code underlies the sound-symbolic use of voice pitch. In Hinton, L., Nichols, J., Ohala, J.J. (eds.), *Sound symbolism* (pp. 325–347). Cambridge: Cambridge University Press.
- Oosterhof, N. N., Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences of the United States of America*, 105, 32, 11087–11092.
- Osgood, C. E. (1964). Semantic differential technique in the comparative study of cultures. *American Anthropologist*, 66, 3, 171–200.
- Perry, T. L., Ohde, R. N., & Ashmead, D. H. (2001). The acoustic bases for gender identification from children’s voices. *The Journal of the Acoustical Society of America*, 109, 6, 2988–2998.
- Ptacek, P. H., & Sander, E. K. (1966). Age recognition from voice. *Journal of Speech and Hearing Research*, 9, 273–277.
- Scherer, K.R. (1986). Vocal affect expression: A review and a model for future research. *Psychological Bulletin*, 99, 143–165.
- Shayan, S., Ozturk, O., & Sicoli, M. A. (2011). The thickness of pitch: Crossmodal

- metaphors in Farsi, Turkish, and Zapotec. *Senses and Society*, 6, 1, 96-105.
- Shochi, T., Aubergé, V., & Rilliard, A. (2005). Because attitudes are social affects, they can be false friends... In J. Tao, T. Tan, & R. W. Picard (eds.) *Affective computing and intelligent interaction: First International Conference, AII 2005, Beijing, China, October 22-24, 2005*, Proceedings (pp. 482-489).
- Teshigawara, M. (2003). *Voices in Japanese animation: A phonetic study of vocal stereotypes of heroes and villains in Japanese culture*. PhD thesis, University of Victoria, Canada. [<http://dspace.library.uvic.ca:8080/handle/1828/361>] (2019年6月12日取得)
- Teshigawara, M. (2011). Voice-based person perception: Two dimensions and their phonetic properties. *Proceedings of the Seventeenth International Congress of Phonetic Sciences, 1974-1977*.
- Teshigawara, M., Amir, N., Amir, O., Milano Wlosko, E., Avivi M. (2009). Perceptions of Japanese anime voices by Hebrew speakers. In: Izdebski, K. (ed.), *Emotions of the Human Voice, Vol. III Culture and Perception* (pp. 189-198). San Diego: Plural Publishing.
- van Bezooijen R. (1988). The relative importance of pronunciation, prosody, and voice quality for the attribution of social status and personality characteristics. In R. van Hout & U. Knops (eds.), *Language attitudes in the Dutch language area* (pp. 85-103). Dordrecht, the Netherlands: Foris.
- Yarmey, A. D. (1993). Stereotypes and recognition memory for faces and voices of good guys and bad guys. *Applied Cognitive Psychology*, 7, 419-431.
- 内田照久 (2009) 「音声の韻律的特徴と話者のパーソナリティ印象の関係性」『音声研究』13 卷 1 号, 17-28.
- 内田照久 (2011) 「音声中の母音の明瞭性が話者の性格印象と話し方の評価に与える影響」『心理学研究』82 卷 5 号, 433-441.
- 籠宮隆之・山住賢司・榎洋一・前川喜久雄 (2007a) 「講演音声の大局的な印象に影響を与える要因」『音声研究』11 卷 2 号, 65-78.

声に関するステレオタイプの解明に向けて

- 籠宮隆之・山住賢司・槇洋一・前川喜久雄 (2007b) 「聴取実験に基づく講演音声の印象評定データの構築とその分析」『社会言語科学』9 卷 2 号, 65-76.
- 神山孝夫 (2008) 『脱・日本語なまり：英語 (+ α) 実践音声学』大阪大学出版会.
- 勅使河原三保子 (2015) 「役割語の音声とその翻訳について」定延利之 (編) 『私たちの日本語研究 一問題のありかと研究のあり方一』, 朝倉書店, pp. 132-136.

謝辞

本研究は筆者が在外研究中に行った研究の成果の一部を基に、JSPS 科研費 18K00547 の助成を受けて行われました。