

顔と声の組み合わせが伝える真の感情に関する認知科学的検討：  
文化を超えた感情翻訳技術の開発に向けて (102103011)  
Emotion perception through the combination of faces and voices

研究代表者

田中 章浩 東京女子大学  
Akihiro Tanaka (Tokyo Woman's Christian University)

研究期間 平成 22 年度～平成 24 年度

概要

異文化の話し手の感情をも「翻訳」できるユニバーサル・コミュニケーション・システムの開発に向けた基礎的知見を得るため、さまざまな感情を表現した顔と声を組み合わせて提示する心理実験を実施した。同様の実験を文化的背景の異なる複数のグループを対象に実施し、それぞれの文化における感情情報の表示規則および解読規則の一端を明らかにした。

1. まえがき

異なる言語を話す人たちの間でコミュニケーションを可能にすることは、社会のグローバル化やボーダーレス化に伴い、ますます重要になってきている。近年の音声翻訳技術は飛躍的に向上し、言語の壁を取り払うことは現実味を帯びてきた。しかし、コミュニケーションでは言語情報に加えて、顔の表情や声の調子等を通して、感情（非言語情報）も同時に伝達されている。感情の表現と解読には普遍的側面と同時に文化依存的側面もあり、異文化間コミュニケーションを妨げる原因となっている。そこで、今後は文化の壁を超えるために、異文化の発話者が表出した感情をも「翻訳」できる新たなコミュニケーション技術が求められる。ここでいう翻訳とは、ある文化における感情表現を解読して、別の文化において理解可能な情報（アバター、ロボット、あるいは実際の人物の表情と音声など）として、話し手の感情情報を提示することをさす。こうした技術の基盤には、さまざまな文化的背景をもつ人間が、顔や声から感情をどのように表出し（表示規則）、どのように解読しているのか（解読規則）についての基礎的知見の体系的蓄積が不可欠である。

本研究では、感情表現の本質は顔や声などの複数の多感覚情報の組み合わせにあると考え、こうしたマルチモーダル情報に基づいた感情の表出と解読の「文法」に迫る。そして、感情情報をも翻訳できるユニバーサル・コミュニケーション・システムの設計指針を得る。具体的には、さまざまな感情を表現した顔と声を組み合わせて提示して、感情を推定させる心理実験を通して、人間の多感覚感情情報処理についての基礎的な知見を得る。そして、同様の実験を複数の文化を対象に実施し、多感覚的な感情表現の表示規則および解読規則の原理（普遍的部分）とパラメータ（文化依存部分）をモデル化する。

2. 研究開発内容及び成果

2.1 マルチモーダル感情発話データベースの作成

人間が他者の顔の表情と音声のパラ言語情報に基づいて、どのように感情を読み取るのかを実験的に検討するためには、発話者が各種感情を顔と声の両方を用いて表現した実験刺激が必要である。しかし、表情と感情音声のそれぞれのデータベースは存在するものの、マルチモーダルな感情表現について基本感情を網羅し、複数の文化で比較可能な形式のデータベースは存在しない。そこで、日本人 21 名およびオランダ人 15 名の協力を得て、顔と声によってさまざまな感情（怒り、喜び、悲しみ、恐れ、驚き、嫌悪）を表現してもらい、その発話映像と音声を収録した。発話スクリプトには、言語的には中立な意味を持つ短文

（例：「これなに？」「そうなんですか」など）を用いた。そのうえで、収録動画を対象とした評定実験を日本人 101 名、オランダ人 130 名を対象に実施し、実験用データセットの選定をおこなった。評定実験では、日蘭の発話サンプルについて、それぞれの文化に属する評価者に、顔のみおよび声のみで提示したときにどのような感情を読み取るかを回答してもらった。評定実験の結果は、混同行列（どの感情表現がどの感情表現として誤認されやすいのかを行列形式で表現した表）として集計した。

表情に関しては、喜び表情と怒り表情の正答率は高い反面、恐怖表情の正答率がとくに日本人において低く、日本人の恐怖表情の認識には大きな困難が伴うことが浮き彫りとなった。ただし、音声に関しては全体的には表情よりも正答率が低いものの、日本人では恐怖表情よりも恐怖音声の正答率が高かったことから、感情の解読において音声は表情と相補的に機能していることが明らかとなり、マルチモーダル感情認識という方向性の妥当性を支持する結果が得られた。



図 1：実験の様子

2.2 基本的感情の表示規則と解読規則の検討

2.1 で開発したデータベースを用いて、日蘭国際比較実験を実施し（図 1）、発話者（収録動画）による基本的感情の表示規則と、観察者（実験参加者）による解読規則の普遍性と文化依存性について検討した。以下に主要な結果の一部を示す。

- ・オランダ人は日本人よりも顔表情認知が正確なのに対し、日本人はオランダ人よりも声感情認知が正確である
- ・自文化（例：発話者も観察者も日本人）と異文化（例：発話者がオランダ人で、観察者は日本人）での判断を比べると、自文化では異文化よりも音声優位となる
- ・顔のみないし声のみでの提示と比べて、顔＋声での提示が常に有利とは限らない
- ・発話者は表情と音声の両方に同じ感情（例：喜び）を込めたつもりでも、観察者側では顔と声からそれぞれ異なる

った感情を知覚するケース（例：表情は喜び、音声は驚き）が相当の割合で存在する

- ・上記のような表現は、感情の組み合わせに応じて顔ないし声のいずれかの感情に吸引されて知覚される（例：日本人では怒り・嫌悪・喜び・驚きは顔優位）

これらの知見に基づくと、感情翻訳技術を開発する際に、顔表情認識と声感情認識の結果を統合するとき、発話者と観察者それぞれの文化的背景や感情の種類に応じて、顔と声の重みづけを適切に変化させることが可能となる。

### 2.3 高次感情の表示規則と解読規則の検討

まず罪悪感や軽蔑などの複雑な高次感情は、複数の基本感情を適切なバランスで組み合わせることによって構成されることを概念レベルで確認した。そのうえで、高次感情の多感覚的表示規則を明らかにするため、表情と音声を組み合わせる高次感情を表す視聴覚表現を作成する実験をおこなった（図 2）。実験の結果、日本人は表情と音声で異なる基本感情を組み合わせる高次感情を表出する傾向があるが（例：恥は表情が悲しみ、音声は恐怖の組み合わせ）、オランダ人は同じ基本感情を組み合わせる高次感情を表出する傾向（例：恥は表情と音声ともに悲しみ）が示された。

次に、このように話し手の顔と声から分散的に表現された感情が、聞き手にどのように解読されているのかを検討した。具体的には、表情と音声の表現する感情が異なる人工的実験刺激（例：表情が喜び、音声は悲しみ）を用いた。実験の結果、日本人ではそれぞれ異なった基本感情の表情と音声を組み合わせるときに高次感情として判断する傾向が確認できた（例：表情が悲しみで音声は恐怖だと、罪悪感として判断する）。



図 2：実験画面

### 2.4 マルチモーダル感情情報処理過程のモデル化

一連の実験結果に基づいて、顔と声による多感覚的な感情の表示規則と解読規則における普遍的側面と文化依存の側面の双方を盛り込んだ、人間の多感覚感情情報処理のモデル化を進めた（図 3）。

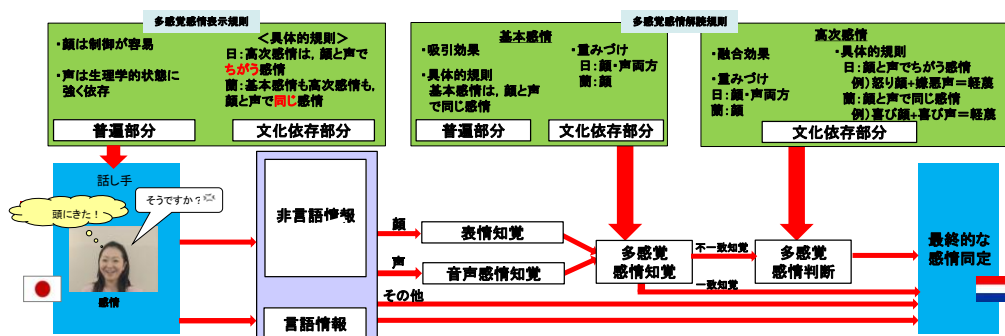


図 3：マルチモーダル感情情報処理過程のモデル

### 3. 今後の研究開発成果の展開及び波及効果創出への取り組み

本研究で得られた知見は、文化の壁を超える感情翻訳技術の開発に直結する基盤となる。今後、顔表情の認識および合成、音声感情の認識および合成、そして音声言語翻訳といった要素技術と本研究成果を組み合わせることで、真に言語と文化の壁を超えるユニバーサル・コミュニケーション技術へと展開していくことが期待される。このような技術は、世界中のあらゆる場所に住むあらゆる文化的背景をもつ人々とほぼ遅延なくコミュニケーションすることを可能にし、新たな交流、より深い相互理解を促進し、地域発の国際事業の発展にもつながる。

本研究成果は、既存の顔と声を組み合わせる多様な感情を表出できるアバターやロボットなどへの応用にもつながる。作成したデータベースは感情認識の工学的研究をはじめ、感情認知の心理学分野でも有効活用できる。今後の産学連携を通して、通信業界、エンタテインメント業界などに成果を波及させたい。

### 4. むすび

本研究から、顔と声に基づく多感覚的な感情の表示規則と解読規則における普遍的側面と文化依存の側面の双方が明らかになった。本研究で得られた知見を感情認識システムに実装することで、異文化コミュニケーションにおける適切な感情伝達の一助となることが期待される。

#### 【誌上発表リスト】

- [1]高木幸子・田部井賢一・Huis In't Veld Lisanne・de Gelder Beatrice・田中章浩、“表情と音声の示す感情が一致していない刺激からの感情知覚—異文化間バーチャル・リアリティ・コミュニケーションへの応用—”、基礎心理学研究（印刷中）
- [2]田中章浩 “顔と声による情動の多感覚コミュニケーション”、認知科学、Vol.18(3)pp.416-427 (2011年9月)
- [3]Tanaka, A., Asakawa, K., & Imai, H. “The change in perceptual synchrony between auditory and visual speech after exposure to asynchronous speech.”、NeuroReport、Vol.22(14) pp.684-688 (2011年9月)

#### 【報道掲載リスト】

- [1] “Le langage émotionnel universel ou culturel?”, Le monde de l'intelligence (フランス)、2013.2.12.
- [2] “55 プラス 話し上手になる(4) 1秒間に6文字の速さで”、朝日新聞（朝刊全国版、家庭欄）、2012.7.2.
- [3] “日本人の感情、顔より声に”、朝日新聞（朝刊全国版、科学面）、2010.12.3.

#### 【本研究開発課題を掲載したホームページ】

<http://akihirotanaka.web.fc2.com>