

# 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—

高木 幸子<sup>a</sup>・田部 井賢一<sup>b</sup>・Elisabeth HUIS IN 'T VELD<sup>c</sup>・  
Beatrice de GELDER<sup>c</sup>・田中 章浩<sup>a\*</sup>  
<sup>a</sup> 東京女子大学・<sup>b</sup> 三重大学・<sup>c</sup> Tilburg University

## Assessment of incongruent emotions in face and voice

Sachiko TAKAGI<sup>a</sup>, Ken-ichi TABEI<sup>b</sup>, Elisabeth HUIS IN 'T VELD<sup>c</sup>,  
Beatrice de GELDER<sup>c</sup>, and Akihiro TANAKA<sup>a\*</sup>

<sup>a</sup> Tokyo Woman's Christian University, <sup>b</sup> Mie University, <sup>c</sup> Tilburg University

Information derived from facial and vocal nonverbal expressions plays an important role in social communication in the real and virtual worlds. In the present study, we investigated cultural differences between Japanese and Dutch participants in the multisensory perception of emotion. We used a face and voice that expressed incongruent emotions as stimuli and conducted two experiments. We presented either the face or voice in Experiment 1, and both the face and voice in Experiment 2. We found that both visual and auditory information were important for Japanese participants judging in-group stimuli, while visual information was more important for other combinations of participants and stimuli. Additionally, we showed that the in-group advantage provided by auditory information was higher in Japanese than Dutch participants. Our findings indicate that audio-visual integration of affective information is modulated by the perceiver's cultural background, and that there are cultural differences between in-group and out-group stimuli.

**Key words:** facial expression, vocal expression, incongruent stimuli, audio-visual integration, virtual reality

### はじめに

ネットワーク技術の普及により、われわれの日常生活においてビデオチャットが新しいコミュニケーションツールとして幅広く利用されるようになった。こうしたツールは相手の表情や身振りを確認しながら会話を進めることができ、遠隔地にいる相手とも対面に近い状況でコミュニケーションをとることを可能にしてくれる。しかし、家族や友人といった既知の相手ではなく、見知らぬ相手とコミュニケーションをとる場合には、ウェブカメラを用いて自分の実画像を伝送することに不安を持つ人も多いただろう。そのため、ユーザ本人の代わりとして、自分の分身であるアバタを用いたコミュニケーションが盛んになっている。一般的に、アバタとはインター

ネット参加者同士のコミュニケーションに用いられるCG（コンピュータ・グラフィックス）やイラストによる参加者の分身であり（湯浅・安村・新田，2003），これを利用することによってユーザは実画像を伝送せずとも表情と音声を用いたコミュニケーションを楽しむことができる。アバタを介したコミュニケーションはユーザ同士のやりとりを促進するとともに（Kurlander, Skelly, & Salesin, 1996; Pesson, 2003; Smith, Farnham, & Drucker, 2000），異文化間コミュニケーションへの参加意欲や親近性を向上させることが、先行研究から明らかになっている（Isbister, Nakanishi, & Ishida, 2000; Koda, 2004）。

社会的動物であるわれわれ人間にとって、コミュニケーションにおいて自分の意図した感情を的確に他者に伝え、他者の意図する感情を正確に読み取ることは非常に重要である。コミュニケーション場面でわれわれが伝え合うメッセージは、言語的情報と非言語的情報に大別することができる。円滑な意思疎通を実現するためには言語的情報は不可欠であるが、特に感情の理解と伝達に

\* Corresponding author. Department of Psychology, Tokyo Woman's Christian University, 2-6-1 Zempukuji, Suginami-ku, Tokyo 167-8585, Japan. E-mail: akihiro@lab.twcu.ac.jp

おいては、表情、声の調子、身振りといった非言語的情報が大きな役割を担っていることが示唆されてきた (Mehrabian, 1968)。

このような知見は、アバタを介したコミュニケーションへのバーチャルリアリティ技術の応用にも利用されている。特に、従来のアバタには非言語的情報が欠落していることが指摘されており (平家・川崎・田中・藤田, 2011), 非言語的情報の付加を試みる研究が数多くおこなわれてきた (平家他, 2011; 石井・渡辺・大崎・伴, 2008; 楠見・米田・小島, 2008; 湯浅・安村・新田, 2003)。こうした研究によって、アバタの自然さやアバタとの対話感が増し、われわれのコミュニケーション形態はより多様で、豊かなものになりつつある。しかしながら、アバタにおける非言語的情報の影響の検討は、アバタの表情の意図的な制御 (楠見他, 2008) や、発話を利用したうなずきの制御 (石井他, 2008), アバタによる自動的な表情表出 (湯浅他, 2003) など、主に表情という視覚的情報に焦点が当てられ、他の感覚情報および複数の感覚情報が同時に与える影響についての検討は少ない。

本研究は、視覚および聴覚情報の両方に焦点を当て、知覚される感情と、受け手の文化的背景および受け手と送り手の文化的背景の組合せによる差を検討するものである。このような視点に基づく基礎的研究は、バーチャルリアリティでの異文化間コミュニケーションをより自然なものに近づけ、活性化させるような知見を提供しうると考えられる。

### 視聴覚情報からの感情知覚

#### 感情の視聴覚統合

非言語的情報からの感情知覚を検討した従来の研究では、主に表情と音声に焦点が当てられ、数多くの研究が単一の感覚モダリティに限定しておこなわれてきた。しかし、実際のコミュニケーション場面では、われわれは複数の感覚モダリティの処理を同時におこない、複雑で豊かな感情のやりとりを実現している。たとえば、表情から手がかりを得ると同時に、声の調子からも手がかりを得て、これらを統合することによって最終的な判断をおこなっているのである。

このように、われわれが他者の感情を理解する際には、表情から得られる視覚情報と声から得られる聴覚情報の間には相互作用が生じることが、多くの先行研究で報告されている (Collignon et al., 2008; de Gelder & Vroomen, 2000; Koizumi et al., 2011; Massaro & Egan, 1996; Tanaka et al., 2010a)。たとえば、de Gelder and Vroomen (2000) は、

異なる感情を示した表情と音声を人工的に組み合わせた刺激を用いて、実験参加者に表情と音声を同時呈示して感情判断を求める実験をおこなった。その結果、表情のみに着目するように教示した場合でも、最終的な感情判断は同時呈示した音声に含まれる感情に影響されることが明らかになった。

アバタを介したチャットコミュニケーションの研究においても、視覚情報と聴覚情報の結びつきが重要であることが示唆されている。宮島・藤田 (2007a) は、音声の基本周波数が快感の強度を反映することに注目して研究をおこなった。彼らの研究では、音声の基本周波数に応じたアバタの笑顔強度の自動制御が、アバタを用いたコミュニケーションの自然さや対話感の主観的評価に及ぼす影響を検討した。アバタ同士のコミュニケーションにおいて、受け手が送り手の音声から期待する表情と送り手のアバタの表情の整合性が高ければ、自然な印象や対話感が増し、会話が促進されると考えたからである。実験の結果、音声の基本周波数を用いてアバタの笑顔強度を制御した場合には、自然さや対話感のみならず、好ましさや被聴取感 (アバタに話を聞いてもらっているという感覚) についての主観的評価が高くなることが明らかになった。つまり、視覚と聴覚、両モダリティが示す感情情報をアバタに取り入れることによって、ユーザとアバタのインタラクションが強化されることが実証された。

#### モダリティ間での感情不一致

宮島・藤田 (2007a) の研究からは、アバタを介したコミュニケーションの自然さや対話感を高める要因の一つとして、表情と音声の整合性が重要であることが示唆された。しかしながら、実際のコミュニケーション場面では必ずしも表情と音声に整合性があるわけではない。特に感情については、表情と音声に整合性がない、すなわち表情と音声の示す感情が一致していない状況はまれではないということが筆者らの研究によって示唆されている (高木・平松・田中, 2011; Takagi, Hiramatsu, Huisin 't Veld, de Gelder, & Tanaka, 2011)。

高木他 (2011) は日本人モデルによる表情音声動画刺激セットを作成する過程で、こうした感情が一致していない状態が自然な、すなわち編集を加えたわけではない非人工的な刺激でも存在していることを見出した。表情音声動画刺激セット作成のための収録では、モデルは基本6感情 (喜び・怒り・悲しみ・嫌悪・驚き・恐怖) のそれぞれを、表情と音声の両方で同時に演技することを求められた。そして、刺激の評価実験として、無音で表情

動画のみを呈示する表情課題と、表情動画なしに音声のみを呈示する音声課題を同一の日本人実験参加者に対して実施し、それぞれの刺激について基本6感情のうちどの感情を知覚したかを回答させた。分析の結果、表情と音声に対する最頻回答が明らかに異なって判断される刺激が全体の1割程度の割合で存在した。たとえば、モデルは表情と音声の両方に喜び感情を込めたつもりでも、表情課題では喜び、音声課題では驚きが最頻回答となるような刺激である。

また、Takagi et al. (2011) は、オランダ人モデルによる表情音声動画刺激セットを作成する過程におけるオランダ人実験参加者による評価実験でも、同様の刺激が全体の2割程度の割合で存在することを見出した。すなわち、感情表出者が表情と音声で同一の感情を表現することを意図して演技したにもかかわらず、それぞれが異なる感情を示していると判断されるような、表情と音声の示す感情が一致していない刺激は、文化によらず一定の割合で存在することが示された。また、日本人刺激では意図した感情が表情にあらわれた刺激は約6割、音声にあらわれた刺激は約4割であった。一方、オランダ人刺激では意図した感情が表情にあらわれた刺激が約9割を占め、表情と音声で感情が一致していない刺激の出現頻度と、その質には文化的背景による差があることが示された。

### 感情の視聴覚統合における文化差

先行研究によって視覚と聴覚で示される感情が一致していない刺激の出現頻度と各モダリティへの意図感情の表出割合に文化差があることが示されたが（高木他, 2011; Takagi et al., 2011）、こうした刺激から知覚される感情にも文化差があることが報告されている。Tanaka et al. (2010a) は、非言語的な感情知覚における多感覚情報の統合様式における文化差を検討するため、顔と声による視聴覚感情知覚の日蘭比較実験をおこなった。日本人とオランダ人を比較対象とした理由は、言語的要因の影響を統制するためであった。たとえば、日本人（日本語母語話者）とアメリカ人（英語母語話者）を比較対象とした場合、日本人は双方の言語をある程度理解できるが、アメリカ人は日本語を理解できないというように、言語理解に非対称性が生じるおそれがある。しかし、日本人とオランダ人の比較であれば言語的要因が排除され、非言語的情報について対称的な比較が可能になる。実験では、喜びと怒りの感情に着目し表情と音声の感情が一致している、または一致していない動画を刺激として用いた。実験において刺激は、表情と音声をそれぞれ単独、および同時という2つの方法で呈示された。表情・音声

同時呈示では、表情と音声の両方の感情を含む動画を呈示し、表情注意課題と音声注意課題の2種類を実施した。表情注意課題では音声を無視して表情から知覚される感情を判断するように、音声注意課題では表情を無視して音声から知覚される感情を判断するように教示した。表情と音声の単独呈示では、無音の動画（表情のみ課題）または映像の伴わない音声（音声のみ課題）のいずれか一方を呈示し、実験参加者に知覚される感情の判断を求めた。2つの呈示方法ともに、判断は「喜び」と「怒り」の二肢強制選択とした。分析の結果、日本人はオランダ人と比べて表情の感情判断時に音声からの影響が大きく、音声の感情判断時には表情からの影響は小さかった。つまり、一貫してオランダ人よりも音声への依存性が高かった。こうした結果が生じた理由として、Tanaka et al. (2010a) は感情知覚の際の中心情報と文脈情報における西洋と東洋での処理方略の差に着目し、日本人はオランダ人と比較して文脈情報である音声に影響を受けやすい可能性を挙げた。同様の傾向がTakagi et al. (2011) でも確認されており、これは感情知覚における多感覚情報の重みづけが受け手の属する文化的背景によって異なることを示唆している。

### 感情知覚における内集団優位性

上述の文化差は、感情知覚における文脈情報の処理過程という観点から、感情の受け手側が属する文化的背景による差について言及していた。しかし、感情の送り手の文化的背景や、送り手側と受け手側の文化的背景の組み合わせによっても感情知覚には差が生じることが知られている。たとえば、表情のみや音声のみといった単独モダリティ情報からの感情知覚の精度は、送り手と受け手が同じ文化圏に属する人物か否かで差があることが先行研究から明らかにされている。このような文化的背景の組み合わせによって生じる文化差は、それぞれの文化特有の感情表現とそれに対する接触頻度という観点から説明されるのが一般的である（Elfenbein & Ambady, 2003）。表情からの感情知覚については、Elfenbein & Ambady (2002) は複数の表情研究のメタ分析を実施し、人間は外集団（異文化）の人々の表情よりも、内集団（同一文化）の人々の表情から、より正確に感情が判断されることを示した。音声からの感情知覚については、Scherer, Banse, & Wallbott (2001) はそれぞれ母語が異なる9カ国の実験参加者に、ドイツ語を母語とする話者による無意味発話を呈示し、知覚される感情について実験的検討をおこなった。その結果、ドイツ語を母語とする実験参加者は、他の言語を母語とする実験参加者よりも

感情知覚の精度が高いことが示された。このような現象は、内集団優位性 (In-group advantage) として知られている。

表情における内集団優位性は、アバタ表情においても生じることが神田・石田 (2006) の研究から明らかになっている。神田・石田 (2006) は、アバタ表情解釈における文化差を検討するために公開Web実験をおこなった。Web実験では、日本人デザイナーによるアバタキャラクターを刺激とし、アジア3カ国 (日本・韓国・中国) と欧米5カ国 (イギリス・フランス・ドイツ・アメリカ・メキシコ) の計8カ国間におけるアバタ表情の解釈内容を比較分析した。その結果、否定的な表情は文化を超えて正しく理解されるが肯定的な表情は文化によって解釈の差が大きいという文化差が得られるとともに、日本人がデザインしたアバタに対しては日本人実験参加者における表情解釈の一致率が高いという内集団優位性がみられることが明らかになった。

重野 (2004) は、表情と音声を単独呈示した場合にはともに内集団優位性がみられることを明らかにしたうえで (重野, 2003), 表情と音声から知覚される感情に及ぼす受け手と送り手の文化的背景の組み合わせの影響を検討した。実験では、基本6感情をあらわした表情と音声を人工的に組み合わせ、両モダリティの感情が一致している、もしくは一致していない視聴覚刺激を使用した。これらを実験参加者に呈示し、知覚される感情について日米比較をおこなった。実験の結果、感情が一致した刺激に対する正答率は日米で差がみられなかったが、一致していない刺激から知覚される感情には受け手と送り手の文化的背景の組み合わせによる差がみられた。日米の実験参加者はともに、表情から知覚される感情を選択する割合は異文化の話者刺激に対して高く、音声から知覚される感情を回答する割合は同一文化の話者刺激に対して高いことが示された。つまり、受け手と同一文化圏に属する人物についての感情判断では音声をより重視する傾向が示された。

重野 (2004) は日米比較によって、基本6感情を網羅した視聴覚情報が統合された場合に知覚される感情における、受け手と送り手の文化的背景の組み合わせの影響を明らかにした。研究結果からは、受け手の文化によらず同一文化集団に対する感情知覚では聴覚モダリティを重視し、異文化集団に対する感情知覚では視覚モダリティを重視する傾向が示唆された。このことから、送り手の属する文化圏によって感情知覚の際に重視するモダリティは異なる傾向があると考えられる。また、日本人とオランダ人の比較でも感情知覚における文化差がみら

れたが (Tanaka et al., 2010a), Tanaka et al. (2010a) は文化によらず感情知覚の際に中心的となる情報はあくまでも表情であり、文脈情報である音声への依存度が文化によって異なると解釈している。

重野 (2003, 2004) の研究は日米間の比較であったため、日本人参加者には日英両言語を理解でき、米国人参加者には英語しか理解できないという非対称性があった。表情と音声の非言語情報による感情知覚プロセスは、発話の言語内容を理解できるかどうかによって異なる可能性を指摘できる。このような可能性を排除するためには、言語内容の理解について対称的な関係にある文化間での比較が有効であると考えられる。

さらに、表情と音声のあらかず感情が一致していない刺激を用いた研究では、表情と音声で同時に感情を表出するのではなく、別々に収録された表情と音声を意図的 (人工的) に組み合わせることによって、各モダリティで感情が一致していない状態を作り出す手法を用いてきた (重野, 2004; Tanaka et al., 2010a)。こうした手法は任意の感情の組み合わせを作成できるという利点をもつ。しかし、高木他 (2011) をふまえると、わざわざ人工的な手法を用いずとも、日常生活でわれわれは意図せずに表情と音声で異なる感情が表出された表現に接しており、そのような生態学的妥当性の高い刺激を用いることにも意義があると考えられる。このように意図せずして表情と音声から異なる感情が表出された刺激を用いて検討した場合にも、意図的に作成した刺激を用いた場合と同様の傾向が得られるのかどうかは明らかではない。

## 目的と仮説

本研究の目的は、意図せずして表情と音声に異なる感情が表出された刺激を用いて、受け手と送り手の文化的背景の組み合わせが感情知覚に及ぼす影響を検討することである。類似の手法を用いて感情の視聴覚統合を検討した Tanaka et al. (2010a) では、感情知覚の際の顔と声という多感覚情報の重みづけにおける文化差に焦点を当てていた。そのため、顔と声の難易度を一定にする必要が生じ、顔にはノイズが加えられていた。また、明確に快-不快に分類可能な喜びと怒りの感情に限定することによって、日本人における聴覚モダリティの優位性を示した。しかし、実際のコミュニケーション場面では顔と声で感情知覚の難易度に差があり、顔と声の示す感情が不一致な状態は少なからず存在し (高木他, 2011; Takagi et al., 2011), そこにはさまざまな感情の組合せが想定される。

そのため、具体的には Tanaka et al. (2010a) の研究に

以下の変更を加え、発展させた。第一に、現実場面に近づけるため、顔へのノイズ付加といったモダリティ間での難易度の調整はしなかった。第二に、感情は快-不快の組み合わせに限定せず、演技の段階で表出されたすべての組み合わせを含めた。第三に、同一文化集団と異文化集団の結果を比較することで、感情の送り手と受け手の文化的背景の組み合わせによって生じる差についても検討した。そして第四に、意図せずして表情と音声に異なる感情が表出された刺激を用いた。

実験は、2つに分けておこなった。まず、実験1では単独モダリティ情報に対する感情知覚に焦点を当てた。先行研究から、単独モダリティ情報からの感情知覚では内集団優位性がみられることが明らかになっている(Elfenbein & Ambady, 2002; Scherer et al., 2001; Scherer, 2003; 重野, 2003)。しかし、一般的に表情のみからの感情知覚は難易度が低いことが知られている(池本・鈴木, 2009; Scherer, 2003)。また、視聴覚モダリティ情報を同時呈示した複雑な状況においても、表情の示す感情は異文化集団の話者に対しても知覚されやすいことが示されている(重野, 2004; Tanaka et al., 2010a)。ゆえに、表情のみからの感情知覚では受け手と送り手の文化的背景の組み合わせが同一である場合の正答率はさほど低下せず、組み合わせが異なる場合でも正答率は高く、感情知覚の精度には大きな違いはみられないと予測される。

一方、音声のみからの感情知覚は難易度が高いことが知られている(池本・鈴木, 2009; Scherer, 2003)。また、視聴覚モダリティ情報を同時呈示した場合には音声の示す感情は同一文化集団の話者に対して知覚されやすいが、異文化集団の話者に対しては知覚されにくいことが示されている(重野, 2004; Tanaka et al., 2010a)。ゆえに、音声のみからの感情知覚では受け手と送り手の文化的背景の組み合わせが同一である場合の方が、組み合わせが異なる場合よりも感情知覚の精度が高いことが予測される。

次に、実験2では視覚と聴覚の両モダリティの感情情報を同時呈示することにより、視聴覚統合におけるモダリティ優位性に焦点を当てた。こうした状況での感情知覚は、音声より表情のほうが感情判断が容易であるため(池本・鈴木, 2009; Scherer, 2003)、受け手の文化的背景によらず主として表情に依存した感情知覚がおこなわれると考えられる。ただし、内集団優位性(Elfenbein & Ambady, 2002; Scherer et al., 2001)の影響により、その程度は送り手と受け手の文化的背景の組み合わせによって異なることが予測される。具体的には、音声による感情知覚では、日本語とオランダ語の類似性が低く感情知覚

が困難なため(Elfenbein & Ambady, 2002; Scherer et al., 2001)、受け手の文化によらず、同一文化集団と比較して異文化集団からの感情知覚では表情への重みづけがより高まると予測される。また、日本人とオランダ人の比較では、オランダ人よりも日本人は聴覚モダリティ情報への依存性が高いことが知られているため(Tanaka et al., 2010a)、同一文化集団および異文化集団の両方に対する感情知覚においてオランダ人よりも日本人は音声への重みづけが高まると予測される。

## 実 験

### 方 法

**実験参加者** 実験1と実験2ともに日本人参加者(以下、日群)は43名、オランダ人参加者(以下、蘭群)は28名であった。日群のうち、20名(男女各10名、平均年齢20.1歳( $SD=1.36$ ))が日本人刺激を、23名(男性11名、女性12名、平均年齢21.1歳( $SD=2.64$ ))がオランダ人刺激を観察した。また、蘭群のうち、13名(男性2名、女性11名、平均年齢20.3歳( $SD=1.18$ ))が日本人刺激を、15名(男性6名、女性9名、平均年齢20.3歳( $SD=1.59$ ))がオランダ人刺激を観察した。

**刺激** 実験1と実験2ともに刺激は、表情音声動画刺激セット作成の過程で収録された日群・蘭群各576(6(感情) $\times$ 8(モデル数) $\times$ 4(発話内容) $\times$ 3(発話回数))の刺激の中から抽出した。4種の発話内容は「そうなんですか」「これなに」「さようなら」「どうなってるの」であり、中立的な意味を持つが、基本6感情のすべての感情を込めて発話される可能性のある短文であった。刺激抽出の際には、各動画の表情と音声を単独で呈示した感情評価実験において、表情と音声に異なる感情をあらわしていると評価され、特定の感情の最頻回答率がそれぞれ50%以上であるという基準を設けた。その結果、日本人刺激56種類(9.7%)、オランダ人刺激105種類(18.2%)が抽出された。56種類の日本人刺激のうち、意図感情が表情に表出されていたのは33種類(58.9%)であり、音声に表出されていたのは23種類(41.1%)であった。また、105種類のオランダ人刺激のうち、意図感情が表情に表出されていたのは98種類(93.3%)であり、音声に表出されていたのは7種類(6.7%)であった。各モダリティにおける異なる感情は意図せず表出されたものであるため、刺激全体に占める感情の組み合わせの割合は一定ではなかったが、日本人刺激とオランダ人刺激を合わせれば意図感情には基本6感情のすべてが含まれていた。また、日本人刺激とオランダ人刺激には、ともに男女各4名ずつの計8名のモデルが含まれていた。

刺激全体に占める表情と音声の感情の組み合わせの内訳を、日本人についてはTable 1に、オランダ人についてはTable 2に示した。Table 1とTable 2の表情と音声の感情は、刺激の評価実験において各刺激の表情と音声に対して最も多く選択された感情である。

**手続き** 実験1では、刺激は表情のみ（以下、V呈示）、音声のみ（以下、A呈示）の2つの方法で日群および蘭群に呈示した。A呈示とV呈示の順はカウンターバランスした。実験2では、刺激を表情・音声同時に呈示した（以下、AV呈示）。実験1と実験2はともに個人形式でおこなわれ、17インチのモニタ（SONY, CPD-E220）とスピーカ（SONY, VGP-SP1）を用いて刺激が呈示された。実験の各試行では、注視点に続いて刺激がモニタおよびスピーカによって呈示され、続いて回答の選択肢として基本6感情が呈示された。基本6感情には、それぞれ1から6までの数字が割りふられていた。いずれの実験でも実験参加者は、モニタに呈示された基本6感情から1つの感情を選択するという形式で、呈示された刺激から知覚される感情を数字キーを使って回答した。回答の感情のキー配列は6通り用意し、実験参加者にランダムに割り当てた。日本人刺激を用いた実験では、実験1のA呈示とV呈示はそれぞれ56試行、実験2も56試行

から構成されており、実験全体の所要時間は約45分であった。オランダ人刺激を用いた実験では、実験1のA呈示とV呈示はそれぞれ105試行、実験2も105試行から構成されており、実験全体の所要時間は約70分であった。

## 実験 1

実験1は、単独モダリティからの感情知覚の傾向について検討するためにおこなわれた。

### 結果と考察

V呈示とA呈示それぞれにおいて、2要因被験者間計画の分散分析を実施した。要因は実験参加者群（日・蘭）×刺激タイプ（同一文化集団・異文化集団）、従属変数は知覚された感情の正答率であり、刺激の評価実験において最頻回答が得られた感情に対する回答率とした。

**V呈示** V呈示の結果をFigure 1に示す。V呈示では、日群の同一文化集団と異文化集団に対する平均正答率はそれぞれ67.9%（SD=10.1）と63.1%（SD=13.0）であり、蘭群での平均正答率は66.7%（SD=11.7）と62.2%（SD=7.3）であった。分散分析の結果、実験参加者群（ $F(1, 67)=0.13$ , n.s.,  $\eta^2=0.00$ ）と刺激タイプ（ $F(1, 67)=2.82$ , n.s.,  $\eta^2=$

Table 1

Combinations of most frequent responses to facial and vocal expressions of Japanese evaluated as expressing different emotions

		Face					
		Happy	Angry	Sad	Disgust	Surprise	Fear
Voice	Happy					1	
	Angry	1			11	1	
	Sad	2			5	4	
	Disgust		5				
	Surprise	4	7				
	Fear			13			2

Note: Values show numbers of stimuli.

Table 2

Combinations of most frequent responses to facial and vocal expressions of Dutch evaluated as expressing different emotions

		Face					
		Happy	Angry	Sad	Disgust	Surprise	Fear
Voice	Happy		3		7	15	4
	Angry	1			3	2	
	Sad				11		2
	Disgust		1				
	Surprise	28	4	2	7		2
	Fear		2	7	2	2	

Note: Values show numbers of stimuli.

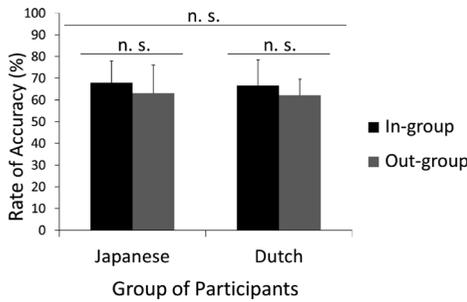


Figure 1. Mean accuracy in the visual presentation in Experiment 1. Error bars show standard deviations.

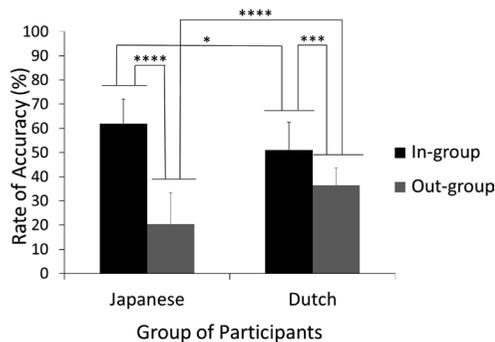


Figure 2. Mean accuracy in the auditory presentation in Experiment 1. Error bars show standard deviations. \* $p < .05$ , \*\*\* $p < .005$ , \*\*\*\* $p < .001$ .

0.04)の主効果と、交互作用 ( $F(1, 67) = 0.00$ , n.s.,  $\eta^2 = 0.00$ )のいずれにも有意差はみられなかった。つまり、表情のみを呈示した場合の正答率に、実験参加者が属する文化的背景および送り手の属する文化的背景による差はみられなかった。

**A呈示** A呈示の結果をFigure 2に示す。A呈示では、日群の同一文化集団と異文化集団に対する平均正答率はそれぞれ61.9% ( $SD = 13.6$ )と20.4% ( $SD = 6.7$ )であり、同様に蘭群での平均正答率は50.9% ( $SD = 7.5$ )と36.4% ( $SD = 8.7$ )であった。分散分析の結果、実験参加者群×刺激タイプの交互作用 ( $F(1, 67) = 31.50$ ,  $p < .001$ ,  $\eta^2 = 0.13$ )と刺激タイプの主効果 ( $F(1, 67) = 134.84$ ,  $p < .001$ ,  $\eta^2 = 0.58$ )がみられた。交互作用について単純主効果検定を実施したところ、実験参加者群ごとで刺激タイプ間に有意差(日群 $p < .001$ ,  $d = 3.96$ ; 蘭群 $p < .001$ ,  $d = 1.80$ )がみられた。つまり、音声のみを呈示した場合、日群・蘭群ともに、異文化集団よりも同一文化集団に対して知覚される感情の正答率の方が高かった。

結果から、表情による感情知覚の精度は、受け手と送

り手の文化的背景の組み合わせによる差がないことが示唆された。一方、音声からは、受け手が日本人でもオランダ人でも異文化集団よりも同一文化集団から感情が読み取りやすいことが示唆された。ただし、同一文化集団からはオランダ人よりも日本人の方が、異文化集団からは日本人よりもオランダの方が正確に感情を読み取る傾向が示唆された。つまり、受け手と送り手の文化的背景の組み合わせが感情知覚に及ぼす影響は呈示モダリティ間で傾向が異なり、音声でのみ内集団優位性と、その程度の強さにおける日蘭差がみられた。

## 実験 2

実験1では単独モダリティによる感情知覚を検討した。実験2では、視覚モダリティと聴覚モダリティを同時呈示することによって、視聴覚統合におけるモダリティ優位性の日蘭文化差を検討した。

## 結果と考察

AV呈示については、2要因被験者間計画の分散分析を実施した。要因は実験参加者群(日・蘭)×刺激タイプ(同一文化集団・異文化集団)であった。従属変数は、実験参加者の表情と一致する感情の回答率と音声と一致する感情の回答率の差分を両者の回答率の和で割った値とし、モダリティ優位性指数とした。モダリティ優位性指数は、実験参加者の表情と音声から知覚される感情の割合について、正の方向に値が大きいほど音声よりも表情から知覚される感情の割合が高く、負の方向に値が大きいほど表情よりも音声から知覚された感情の割合が高かったことを示す。また、0に値が近いほど両モダリティから知覚される感情の割合に差がなかったことを示す。

AV呈示の結果をFigure 3に示す。AV呈示では、同一文化集団と異文化集団に対するモダリティ優位性指数は、日群ではそれぞれ0.03 ( $SD = 0.27$ )と0.65 ( $SD = 0.11$ )であり、蘭群では0.52 ( $SD = 0.20$ )と0.42 ( $SD = 0.10$ )であった。分散分析の結果、実験参加者群×刺激タイプの交互作用 ( $F(1, 67) = 61.36$ ,  $p < .001$ ,  $\eta^2 = 0.40$ )と、実験参加者群 ( $F(1, 67) = 7.80$ ,  $p < .05$ ,  $\eta^2 = 0.05$ )および刺激タイプ ( $F(1, 67) = 32.95$ ,  $p < .001$ ,  $\eta^2 = 0.19$ )の主効果がみられた。交互作用について単純主効果検定を実施したところ、日本人実験参加者の群において刺激タイプ間に有意差がみられた ( $p < .001$ ,  $d = 3.20$ )。この結果は、日群では同一文化集団よりも異文化集団に対して表情に示された感情を知覚する割合が高かったことを示している。

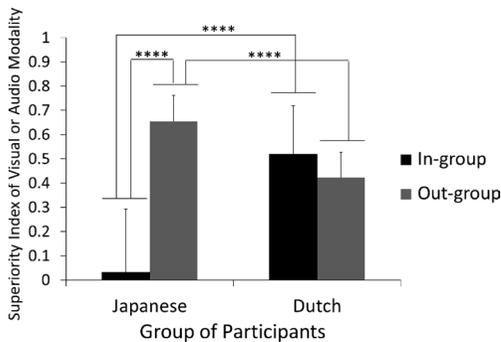


Figure 3. Modality superiority index. Positive values show visual superiority, whereas negative values show auditory superiority.  
\*\*\*\*  $p < .001$ .

## 総合考察

### 実験1の結果

実験1では、表情と音声とを単独呈示した場合の受け手と送り手の文化的背景の組み合わせが感情知覚に及ぼす影響を検討した。

音声のみからの感情知覚には、内集団優位性と、その程度に群間差がみられた。内集団優位性がみられたことは、先行研究 (Elfenbein & Ambady, 2002; Scherer et al., 2001) の結果を支持していた。今回の実験は Tanaka et al. (2010a) に倣い、互いに言語理解が不可能である日蘭での比較とした。先行研究より、音声による感情知覚の精度には、言語の類似性が影響を与えることが明らかになっている (Elfenbein & Ambady, 2002; Scherer et al., 2001)。音声による感情表現は、ピッチや抑揚などの音響特徴の変化によるところが大きく、それらをどのように変化させるかは言語によって異なり、類似性が高い言語間では変化様式にも共通部分が多いからである。ゆえに発話が母語でなされていた同一文化集団に対する感情知覚の精度がより高くなったと考えられる。

さらに、内集団優位性の程度にも受け手の属する文化的背景による差があるという知見が得られた。内集団優位性は群を問わずみられたが、その程度は日本人の方がより強かった。音声における内集団優位性については、Takagi et al. (2011) の研究では強さの程度は日本人よりもオランダ人の方が強いという結果が報告されている。Takagi et al. (2011) と本実験のちがいは、刺激に含まれる感情の割合であった。Takagi et al. (2011) では、収録した576種類の刺激のすべての音声を呈示しており、それらの刺激には基本6感情すべてがほぼ均等に含まれていた。しかし、本実験ではモダリティ間で意図せずして

異なる感情が表出された刺激のみを用いたため、日本人刺激とオランダ人刺激に含まれる基本6感情の割合が均等ではなかった。ゆえに、本実験で得られた知見が先行研究と異なっていた点は、呈示した刺激全体に占める感情の割合のちがいに起因すると考えられる。ただし、Takagi et al. (2011) の研究は収録した刺激の評価を目的としたものであり、基本6感情の表出が不明瞭な刺激も含まれていた。内集団優位性の程度における文化差については、基本6感情が明確に表現され、かつ刺激全体に各感情が均等に含まれた音声刺激を用いた実験によって、さらに検討が必要であろう。また、本研究は感情の送り手と受け手の文化的背景の組合せを厳密に統制し、受け手が同一文化集団と異文化集団の相対的な比較によって感情を知覚することを避けるため、すべての要因において被験者間計画を採用した。そのため、内集団優位性の程度について実験参加者群ごとに同一文化集団と異文化集団に対する感情知覚の正答率の差を算出し、直接的な比較をおこなうことは不可能であった。被験者内計画での実験をおこない、内集団優位性の程度をより詳細に検討することも今後の課題である。

なお、音声のみでは難易度の高さ (池本・鈴木, 2009; Scherer, 2003) と聴覚モダリティへの依存性の違い (Tanaka et al., 2010a) から、送り手が日本人であるかオランダ人であるかによらず、日本人の受け手のほうが、オランダ人よりも精度が高いと予測していた。しかし、実験の結果はこの予測を支持せず、実験参加者群の主効果は有意ではなかった。この結果は高木他 (2011) と異なるものである。このような違いが生じた理由としては、上述の内集団優位性の違いと同様に、先行研究と本研究で呈示した刺激全体に占める感情の割合の違いに起因する可能性を指摘できる。

以上の通り、音声のみからの感情知覚には内集団優位性がみられた。表情のみからの感情知覚にも文化的背景の影響がみられ、内集団優位性が生じるとする先行研究もあるが (Elfenbein & Ambady, 2002; 重野, 2003; Takagi et al., 2011)、本実験の結果では文化的背景の影響はみられなかった。その理由として、表情の普遍性の高さが挙げられる。先行研究から、一般的に音声よりも表情による感情知覚の方が普遍性が高いことが明らかになっている (Ekman & Friesen, 1971; Ekman, Friesen, & Ellsworth, 1982; Russell, 1994; Russell, Lewicka, & Niit, 1989)。それゆえ、感情知覚の難易度は音声よりも表情の方が低くなり、実験参加者群の属する文化的背景および内集団・外集団という受け手と送り手の文化的背景の組合せによる知覚の差がみられなかった可能性がある。

## 実験2の結果

実験2では表情と音声を同時呈示することによって、視聴覚統合におけるモダリティ優位性の日蘭文化差を検討した。表情と音声を同時呈示した場合には、同一文化集団に対しては、日群では表情ないしは音声にあらわれた感情のどちらも知覚される傾向があった。一方、蘭群では表情のみにあらわれた感情が知覚される傾向が強かった。つまり、いずれもどちらかのモダリティの感情を回答したが、日群では表情と音声から知覚される感情の割合に差がないのに対し、蘭群では表情のみから感情が知覚されていた。こうした違いはモダリティ優位性によって説明できる。先に述べたように、喜びと怒りの感情に着目した場合にはオランダ人よりも日本人は感情読み取りの際に聴覚モダリティへの依存が強く (Tanaka et al., 2010a)、基本6感情を網羅した今回の実験刺激をモダリティ情報ごとに呈示した研究でも日本人はオランダ人と比較して音声からの感情知覚の精度が高いことが示されている (Takagi et al., 2011)。ただし、この傾向はあくまでも日本人とオランダ人の間で相対的にみられたものである。Tanaka et al. (2010a) の実験では、顔にノイズを付加してモダリティ間での感情知覚の難易度を調整していた。しかし、モダリティ間の難易度を調整せずに視聴覚刺激を呈示した場合には、日本人の場合でも視覚モダリティ情報の影響が非常に強いことが明らかになっている (Tanaka et al., 2010b)。とはいえ、視覚モダリティ情報である表情に強く影響を受けてもなお、相対的にはオランダ人よりも日本人の方が聴覚モダリティへの依存度が高い。したがって、表情と音声と同時に呈示された場合の同一文化集団に対する感情知覚では、オランダ人の感情読み取りはおもに表情に委ねられるが、日本人では音声からも感情が知覚されるという結果になったと解釈できる。

また、異文化集団に対しては、いずれの群も表情にあらわれた感情が知覚される傾向が示された。つまり、異文化集団に対する感情知覚においては、実験参加者の文化的背景によらず、おもに表情から感情が知覚される傾向が強いと解釈できる。これは、単独モダリティからの感情知覚について検討した実験1の結果とも整合する。実験1では受け手と送り手の文化の影響は表情による感情知覚ではみられないが、音声では内集団優位性がみられることが明らかになった。内集団優位性がみられるということは、外集団の音声に対する判断が相対的に不正確であるということである。表情と音声を同時呈示した場合の感情知覚にもこの傾向が反映され、外集団に対する感情知覚では、音声の判断が困難であるため、おもに表情に依存した感情知覚がおこなわれたと考えられる。

## 感情知覚における文化差のアバタ制御への応用の可能性

近年、バーチャルリアリティ環境におけるアバタはますます擬人化され、アバタを介したコミュニケーションはより豊かで複雑になってきている。たとえば、アバタの動きにユーザ本人の表情や音声の変化を自動で反映させ、制御することが可能である。表情であれば、ユーザの目尻や口角といった表情生成に重要な特徴の変化を抽出し、アバタの表情に反映させる。音声であれば、会話中のユーザの音声におけるピッチや抑揚といった特徴を抽出してアバタ音声に反映させる。また、これらを組み合わせ、ユーザの実際の表情をアバタの表情に反映させるとともに、ユーザの会話音声の特徴をさらにアバタの表情に反映させ、表情の強度を自動で制御することも可能である。本研究によって得られた知見は、このようなアバタを介したコミュニケーション技術に以下のように応用することができる。

予備実験の結果、日本人とオランダ人ではモダリティ間で感情が一致していない刺激の出現頻度はオランダ人の方が多く、意図した感情が表情に表出されるのは日本人では6割程度だが、オランダ人では9割程度であった。これを踏まえると、以下のような解釈が可能になる。送り手のアバタのユーザがオランダ人の場合には、受け手の文化的背景によらず表情のみから感情が知覚される。そのため、音声の基本周波数や抑揚といったパラメータを指標に送り手アバタの表情の変化や強度を自動制御すると、意図した感情が伝わらない恐れがある。また、同様のやり方で自動制御をおこなうとすると、送り手のアバタのユーザが日本人の場合には、受け手も日本人ならば両モダリティから感情を知覚する。そのため、受け手が日本人ならば齟齬は生じないが、受け手がオランダ人ならば表情のみから感情を知覚するために、意図した感情が伝わりにくくなる。ゆえに、送り手がオランダ人の場合にはつねに表情パラメータを指標に音声パラメータを制御し、送り手が日本人の場合には受け手の文化的背景を考慮したうえで指標とするパラメータを決定すると、円滑なコミュニケーションが達成されるだろう。

本実験で用いたのは、意図せずして表情と音声の示す感情が異なる刺激であった。しかし、近年利用されているアバタチャットシステムでは、アバタの各モダリティに表出される感情の変化はユーザの意図的な手動操作によって作り出されるのが一般的である (宮島・藤田, 2007b)。また、目的こそ異なるものの表情と音声で異なる感情を示す刺激を用いた視聴覚統合を検討した先行研究においても、2つのモダリティ間の感情が一致していない状態を人工的に作り出し、刺激として用いることが

多い。この点に着目し、高木・田部井・田中(2012)は、モダリティ間で異なる感情の表出が編集加工による意図的な刺激と、編集によらず意図的ではない刺激から知覚される感情が等価であるかを検討する実験をおこなった。実験では、表情と音声で一致していない感情の表出が意図的ではない(未編集)な動画刺激と、意図的(編集加工)な動画刺激を用い、それらを表情のみ、音声のみ、表情・音声同時の3つの呈示方法で日本人実験参加者に呈示した。実験の結果、感情知覚の傾向は、刺激における一致していない感情の表出が意図的か否かによる差はみられなかった。特に、刺激の種類によらず、表情と音声を同時呈示した場合には各モダリティに示された感情のどちらかと一致する感情が知覚される傾向がみられた。つまり、一致していない感情の表出が意図的な刺激と意図的ではない刺激から知覚される感情は等価であることが示された。ゆえに、本研究で得られた知見は、各モダリティへの感情表出が意図的な操作によってなされるアバタを介したコミュニケーションにも十分に応用可能であると考えられる。

### おわりに

本研究では、視聴覚モダリティそれぞれの示す感情が一致していない刺激から知覚される感情と、受け手と送り手の文化的背景の組み合わせが感情知覚に及ぼす影響を検討することを目的に、演技の段階で意図せずして表情と音声に一致していない感情が表出された刺激を用いて、日蘭比較をおこなった。実験の結果、感情の視聴覚統合において日蘭で文化差がみられた。特に、一般的には感情知覚の手がかりとして聴覚モダリティよりも視覚モダリティから得られる情報が優位であるが、日本人では同一文化集団に対する感情知覚では視覚モダリティのみならず聴覚モダリティの影響も強く受けることが示された。本研究で得られた知見は、バーチャルリアリティ環境におけるアバタの自動制御に応用することができる。バーチャルリアリティ環境における異文化コミュニケーションの齟齬を招かないためには、文化ごとの傾向を考慮した表情と音声のパラメータ制御が必要であろう。具体的には、アバタのユーザが日本人の場合には受け手の文化を考慮して表情と音声どちらを指標とすることを切り換えるべきだが、ユーザがオランダ人の場合にはつねに表情を指標として音声を自動制御する必要がある。それによって、より円滑なコミュニケーションが実現されるだろう。こうした知見を、より自然で、文化的齟齬の少ない、バーチャルリアリティ技術を駆使したコミュニケーションシステムの実現のために提供すること

を目指したい。

### 謝 辞

本研究は、総務省SCOPE若手ICT研究者育成型研究開発(10210311)、およびJSPS科研費若手研究(A)(24680030)の支援を受けました。

### 引用文献

- Collignon, O., Girard, S., Gosselin, F., Roy, S., Saint-Amour, D., Lassonde, M., & Lepore, F. (2008). Audio-visual integration of emotion expression. *Brain Research*, **1242**, 126-135.
- de Gelder, B., & Vroomen, J. (2000). The perception of emotions by ear and by eye. *Cognition & Emotion*, **14**, 289-311.
- Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology*, **17**, 124-129.
- Ekman, P., Friesen, W. V., & Ellsworth, P. (1982). What are the relative contributions of facial behavior and contextual information to the judgement of emotion? In P. Ekman (Ed.), *Emotion in the human face*. 2<sup>nd</sup> ed. Cambridge: Cambridge University Press. pp. 111-127.
- Elfenbein, H. A., & Ambady, N. (2002). On the universality and cultural specificity of emotion recognition: A meta-analysis. *Psychological Bulletin*, **128**, 203-235.
- Elfenbein, H. A., & Ambady, N. (2003). Cultural similarity's consequences: A distance perspective on cross-cultural differences in emotion recognition. *Journal of Cross-Cultural Psychology*, **34**(1), 92-110.
- 平家雅之・川崎裕達・田中貴紘・藤田欣也(2011). 類似度と好感度を両立する個人化アバタのデフォルメ法の検討 ヒューマンインターフェース学会論文誌, **13**(3), 67-78.
- (Heike, M., Kawasaki, H., Tanaka, T., & Fujita, K. (2011). Study on deformation rule of personalized-avatar for producing sense of affinity. *Journal of Human Interface Society: Human Interface*, **13**(3), 67-78.)
- 池本真知子・鈴木直人(2009). 感情判別における声質の影響—単音節, 短文を用いて— 感情心理学研究, **16**(3), 209-219.
- (Ikemoto, M., & Suzuki, N. (2009). The effect of voice quality in discriminating between emotions based on a listening task using vowels and sentences. *The Japanese Journal of Research on Emotions*, **16**(3), 209-219.)
- Isbister, K., Nakanishi, H., & Ishida, T. (2000). Helper agent: Designing and assistant for human-human interaction in a virtual meeting space. *Proceedings of Human Factors in Computing Systems (CHI2000)*. ACM Press. 57-64.
- 石井 裕・渡辺富夫・大崎浩司・伴 好弘(2008). 意識的な手指動作入力による身体的アバタの操作特性分析 ヒューマンインターフェース学会論文誌, **10**(4), 85-94.
- (Ishii, Y., Watanabe, T., Osaki, H., & Ban, Y. (2008). Analysis of embodied avatar manipulation by conscious hand motion input. *Journal of Human Interface Society: Human Interface*, **10**(4), 85-94.)

- Koda, T. (2004). Interpretation of expressive characters in an intercultural communication. *8<sup>th</sup> International Conference on Knowledge-Based Intelligent Information & Engineering Systems (KES2004)*. LNAI3214, Part II, Springer-Verlag, Berlin. 862-868.
- 神田智子・石田 亨 (2006). アバタ表情解釈の異文化間比較情報処理学会論文誌, **47**(3), 731-738.  
(Koda, T., & Ishida, T. (2006). Cross-cultural comparison of interpretation of avatars' facial expressions. *Journal of Information Processing Society of Japan*, **47**(3), 731-738.)
- Koizumi, A., Tanaka, A., Imai, H., Hiramatsu, S., Hiramoto, E., Sato, T., & de Gelder, B. (2011). The effects of anxiety on the interpretation of emotion in the face-voice pairs. *Experimental Brain Research*, **213**, 275-282.
- Kurlander, D., Skelly, T., & Salesin, D. (1996). *Comic Chat. Proceedings of Computer Graphics and Interactive Techniques*. ACM Press, New York. 225-236.
- 楠見 孝・米田英嗣・小島隆次 (2008). アバターの感情表出機能によるマルチユーザ仮想空間コミュニケーションシステムの改良 日本教育工学会論文誌, **31**(4), 415-424.  
(Kusumi, T., Komeda, H., & Kojima, T. Improving communication in 3D-MUD (Multi User Dungeon) by using avatar's facial expression features. *Japan Journal of Educational Technology*, **31**(4), 415-424.)
- Massaro, W., & Egan, P. B. (1996). Perceiving affect from the voice and the face, *Psychonomic Bulletin & Review*, **3**, 215-221.
- Mehrabian, A. (1968). Communication without words. *Psychological Today*, **2**, 53-55.
- 宮島俊光・藤田欣也 (2007a). 音声チャットシステムにおける基本周波数と音圧を利用したアバタ表情制御法 ヒューマンインターフェース学会論文誌, **9**(4), 85-94.  
(Miyajima, T., & Fujita, K. (2007a). Control of avater's facial expression using fundamental frequency and sound pressure in Multi-user voice chat system. *Journal of Human Interface Society: Human Interface*, **9**(4), 85-94.)
- 宮島俊光・藤田欣也 (2007b). 共有仮想空間アバタ音声チャットシステムにおける音声情報を利用した表情制御 信学技報, HCS2007-6, 31-34.  
(Miyajima, T., & Fujita, K. (2007b). Control of avatar's facial expression using voice information in virtual-space voice chat system. *IEICE Technical Report*, HCS2007-6, 31-34.)
- Pesson, P. (2003). ExMS: An animated and avatar based messaging system for expressive peer communication. *Proceedings of GROUP*, ACM Press, New York. 31-39.
- Russell, J. A. (1994). Is there universal recognition of emotion from facial expression?: A review of the cross-cultural studies. *Psychological Bulletin*, **115**, 102-141.
- Russell, J. A., Lewicka, M., & Niit, T. (1989). A cross-cultural study of a circumplex model of affect. *Journal of Personality and Social Psychology*, **57**(5), 848-856.
- Scherer, K. R. (2003). Vocal communication of emotion: A review of research paradigms. *Speech Communication*, **40**, 227-256.
- Scherer, K. R., Banse, R., & Wallbott, H. G. (2001). Emotion influences from vocal expression correlate across languages and Cultures. *Journal of Cross-Cultural Psychology*, **32**, 76-92.
- 重野 純 (2003). 音声と表情による感情の認知一日米間の比較— 青山心理学研究, **3**, 1-8.  
(Shigeno, S. (2003). Recognition of emotion transmitted by vocal and facial expression: Comparison between the Japanese and the American. *The AGU Journal of Psychology*, **3**, 1-8.)
- 重野 純 (2004). 音声と表情の間で矛盾している感情の認知一日米間の比較— 青山心理学研究, **4**, 17-26.  
(Shigeno, S. (2004). Recognition of emotion transmitted by the incongruent expression between voice and face: Comparison between the Japanese and the American. *The AGU Journal of Psychology*, **4**, 17-26.)
- Smith, M. A., Farnham, S. D., & Drucker, S. M. (2000). The social life of small graphical chat spaces. *Proceedings of CHI*. ACM Press, New York. 462-469.
- 高木幸子・平松沙織・田中章浩 (2011). 日本人の顔と声による感情表現の収録とその評価 信学技報, HIP2011-50, 51-56.  
(Takagi, S., Hiramatsu, S., & Tanaka, A. (2011). Recording and validation of audiovisual expressions by Japanese faces and voices. *IEICE Technical Report*, HIP2011-50, 51-56.)
- Takagi, S., Hiramatsu, S., Huis in 't Veld, E. M. J., de Gelder, B., & Tanaka, A. (2011). Recording and validation of audiovisual expressions by faces and voices. *i-Perception*, **2**(8), 775.
- 高木幸子・田部井賢一・田中章浩 (2012). 顔と声のあらかず感情が不一致な刺激に対する感情判断および印象評定 日本音響学会聴覚研究会資料, **42**, 87-92.  
(Takagi, S., Tabei, K., & Tanaka, A. (2012). Emotion judgment and impression rating of audiovisual stimuli representing different emotions from faces and voices. *Transaction on Technical Committee of Psychological and Physiological Acoustics*, **42**, 87-92.)
- Tanaka, A., Koizumi, A., Imai, H., Hiramatsu, S., Hiramoto, E., & de Gelder, B. (2010a). I feel your voice: Cultural differences in the multisensory perception of emotion. *Psychological Science*, **21**, 1259-1262.
- Tanaka, A., Koizumi, A., Imai, H., Hiramatsu, S., Hiramoto, E., & de Gelder, B. (2010b). Cross-cultural differences in the multisensory perception of emotion. *Proceedings of the International Conference on Auditory-Visual Speech Processing 2010*, 49-53.
- 湯浅将英・安村禎明・新田克己 (2003). ベイジアンネットを用いた交渉エージェントの表情表出 情報処理学会論文誌, **44**(11), 1-9.  
(Yuasa, M., Yasumura, Y., & Nitta, K. (2003). A method of generating facial expressions for negotiation agents by using bayesian network. *Journal of Information Processing Society of Japan*, **44**(11), 1-9.)