

表情と音声の情動知覚における視聴覚相互作用： 情動判断と反応時間の分析

Audio-visual Interaction in the Emotion Perception of Face and Voice: Analyses of Emotion Judgment and Reaction Time

正田 悠^{†‡}, 新田 晴[§], 鈴木 紀子[¶], 岸本 和香[†], 阪田 真己子[†]
Haruka Shoda, Haru Nitta, Noriko Suzuki, Waka Kishimoto, Mamiko Sakata

[†] 同志社大学文化情報学部, [‡] 日本学術振興会特別研究員

[§] 同志社大学大学院文化情報学研究科, [¶] 帝塚山大学経営学部

Faculty of Culture and Information Science, Doshisha University

hshoda@mail.doshisha.ac.jp

Abstract

We communicate emotions via facial and vocal expressions, which are generally congruent with each other, e.g., smiling with high-tone voice in case of happiness. On the other hand, these expressions are sometimes conflicting, e.g., smiling with intimidating voice. In the present study, we investigated how people perceive such emotionally congruent and conflicting expressions to understand the mechanism of cross-modal perception. As experimental stimuli, a male actor expressed four emotions (i.e., happy, angry, sad, relaxed), which were selected from Russell (1980)'s circumplex model (positive/negative, high-activity/low-activity), for a meaningless word, "nire-nako." In the experiment, 36 students were exposed to a total of 16 combined face-and-voice movies, each of which was rated on a visual analogue scale in terms of either valence (positive/negative) or activity (high/low). For each stimulus, they also chose from four emotions (i.e., happiness, anger, sadness, relaxation) what emotion the actor expressed. Results showed that observers judged the valence generally in accordance with the facial expression, so that the negative "face" led to faster response to observers. Any perceptual bias between face and voice was not shown in the activity ratings, whereas people perceiving higher activity generated faster responses regardless of the combination of face and voice. The dimensional approach in the present study suggests that the visual dominance for multi-modal stimuli can only be applicable to the "valence" dimension but not to the "activity" dimension. Replicability by using other expressive media (e.g., music performance) and/or other actors awaits future study.

Keywords — Emotion, Audio-visual Interaction, Face, Voice, Circumplex Model, Reaction Time

1. はじめに

日常の対話場面において、我々は自身が持つ文化的背景のもと、複数のモダリティから統合的に感情を判断することで円滑にコミュニケーションを行っている(田中, 2011)。その中でも、視覚情報(e.g., 目や口の動き)と聴覚情報(e.g., 声の高さや大きさ、抑揚)は話し手と受け手を媒介する情報として重要な役割を果たしている(大坊, 1998; Mehrabian, 1972)。本研究では、顔と声によって表現される情動が一致しているときと(e.g., 笑顔で大きくハキハキした声で話す)、両者が表現する情動が異なるときで(e.g., 笑顔で威圧的な声を出す)、受け手が知覚する情動がいかに異なるのかを調べた。

視聴覚の情報が統合されることで、表現の主体から受け手への情動のコミュニケーションにいかなる影響があるのかが調べられてきた(e.g., 荒井・井上・小野・板倉・北崎, 2011; Ohgushi & Hattori, 1996; 田中, 2011; Tanaka, Koizumi, Imai, Hiramatsu, Hiramoto, & de Gelder, 2010; 渡辺・望月, 2004)。話し手が顔と音声で異なる情動を表出したとき、付加した情報に影響されて情動が知覚されるという知見がある一方で(荒井ら, 2011; Tanaka et al., 2010)、主として「表情」で表された情動が知覚されるという視覚優位性に関する報告もある(e.g., 「悲しみ」の表情に「喜び」の声が付随するとき、受け手は「悲しみ」の情動を知覚する)(渡辺・望月, 2004)。また、特定の言語的意味を持たないヴォカリーズ¹の演奏においても、音楽家が意図した情動(e.g., 喜び、悲しみ、怒り、怖れ、無表情)が顔表情と音声とが合わされたときに聴き手によく伝わることを示されている(Ohgushi & Hattori, 1996)。

以上の先行研究では、エクマンの基本情動説(Ekman & Friesen, 1975)に代表されるように、情動を「喜び」、「悲しみ」といった個々の具体的な情動と

¹ 歌詞のない声楽曲のこと

して“カテゴリー的”に扱っている。一方で、個々の情動を離散的なものとして捉えるのではなく、情動を2次元あるいは3次元上に集約し、各次元において連続的に変化するものとして扱うこともできる(e.g., Russell, 1980, “情動次元説”)。Russell (1980)は、一般にヒトが表出・知覚する情動は二次元(以下、「快・不快」「活動性」)で表すことができ、この二次元は独立なものであることを示している(Russell, 1980, 図1)。本研究では、異なる情動を表現した顔表情と音声を視聴した際に、観察者がどのような情動を受け取るのかを、カテゴリー判断と情動次元説に基づいて調べることで、顔と声のクロスモーダル知覚のメカニズムを明らかにすることを試みた。これにより、これまで示されてきた視聴覚統合の知見(e.g., Tanaka et al., 2010; 渡辺・望月, 2004)が「快・不快」「活動性」の両方について適用可能なものなのか、あるいはどちらか一方の影響を受けたものであったのかを明らかにする。

こうした顔と声の視聴覚統合のメカニズムを明らかにするためのもう一つのアプローチとして、刺激が呈示されてから視聴者が情動を判断するまでの反応時間を測定することも有効である。渡辺・望月(2004)は、表情と音声で表される情動が不一致の場合、両者が一致している場合よりも情動判断に至るまでに時間がかかることを示している。本研究はこれを次元的に調べ、「快・不快」および「活動性」の情動判断の過程に顔と声の情動の組み合わせが及ぼす影響を明らかにする。

本研究では、情動を「快・不快」ならびに「活動

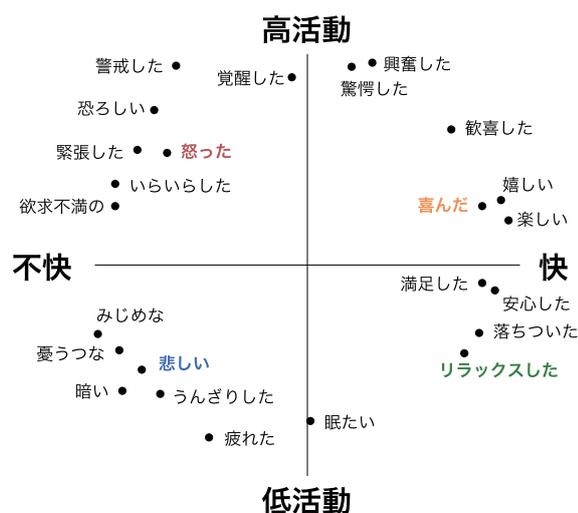


図1. Russell (1980)の感情表現語の二次元モデル。日本語訳は大串(2008)による。

性」からなる2次元上で変化するものと捉え、各次元での情動判断とその反応時間に顔と声の情動の組み合わせが及ぼす影響を調べた。音声よりも表情の方が「快・不快」情報を判断しやすく(高木・平松・田中, 2011)、ヒトはネガティブな情報に対して敏感(渡辺・望月, 2004)な傾向があることから、不快表情における音声の影響は弱いと予測され、また、不快情動に対する反応時間が早いと予測される。

2. 方法

2.1 刺激の作成とその妥当性の確認

2.1.1 顔表情および音声の収録

演劇経験8年の男性1人に、無意味音節「これなに」²を二次元情動モデル(図1)上の各象限を代表する4つの情動カテゴリー(喜び、怒り、悲しみ、リラックス)を表現するように話し分けてもらった。その際、Ekman and Friesen (1975)のアクション・ユニットに基づいて表情を作ってもらった。表情はデジタルビデオカメラ(HDR-XR550, Sony)で収録し、音声はワイヤレスマイク(ECM-AW3, Sony)を介してオーディオレコーダー(R-44, Roland)に録音した。

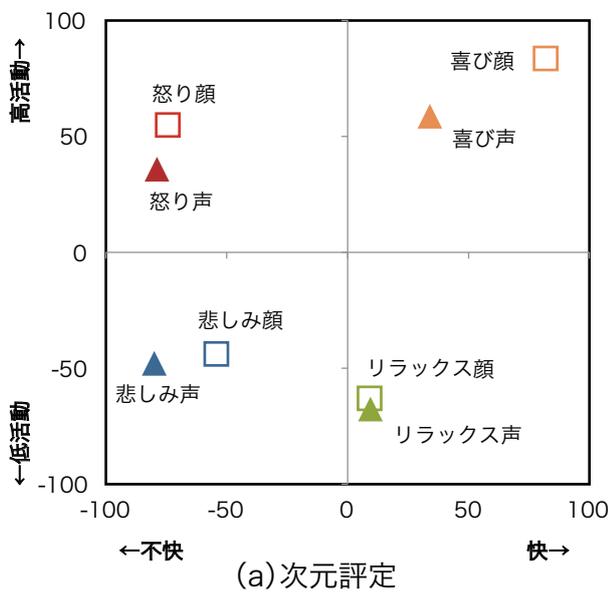
2.1.2 刺激の妥当性の確認(妥当性確認実験)

刺激の妥当性を確かめるために、収録された顔および音声から演者の意図した通りの情動が受け取られるのかを調べる実験を行った。18~34歳($M = 21.58$, $SD = 3.23$)の大学生、大学院生および大学教員24人(男性10人、女性14人)が実験に参加した。参加者に映像および音声をそれぞれ呈示し、演者が表現した「快・不快」および「活動性」の程度を評定してもらった。モダリティ(i.e., 顔, 声)および尺度(i.e., 快・不快, 活動性)の順序はカウンターバランスし、各試行内の4刺激(i.e., 喜び、怒り、悲しみ、リラックス)はランダムに呈示した。刺激呈示および回答収集はSuperLab 4.5 (Cedrus)を用いて行い、刺激呈示後ただちに表示される視覚的アナログスケール上で、演者が表現した「不快(-100)―快(100)」ならびに「活動性低(-100)―活動性高(100)」の程度を直線上をクリックすることで回答してもらった(“評定尺度”)。評定尺度課題の後、もう一度映像と音声をそれぞれ呈示し、それぞれの顔あるいは声が「喜び」、「怒り」、「悲しみ」、「リラックス」のうちどれを表現したものであったのかを選択するよう求めた(“

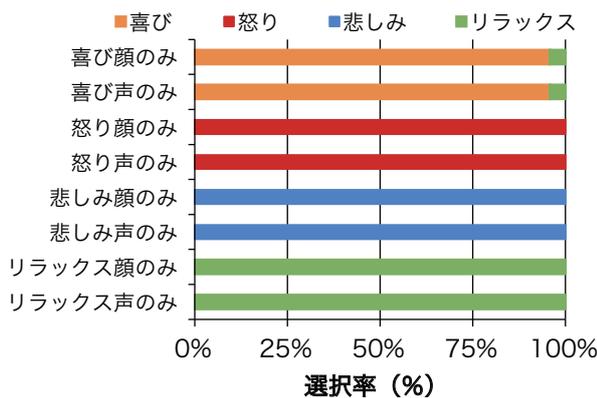
²高木ら(2011)が用いた有意味音節「これなに」のアナグラムである。

情動カテゴリー判断”)。

評定尺度課題における「不快—快」(x 軸)、「低活動—高活動」(y 軸)の各平均値を2次元平面上にプロットした(図2a)。顔、声ともに、喜びが第1象限、怒りが第2象限、悲しみが第3象限、リラックスが第4象限に分類され、演者が意図した通りに参加者が情動を受け取ったことがわかった。同様に情動カテゴリー判断(図2b)についても、怒り、悲しみ、リラックスについては100%の選択率、喜びについても顔と声のそれぞれで「リラックス」と誤った参加者が1人いたのみであった。以上から、95%以上の選択率で演者が意図した通りの情動カテゴリーが受け取られることが確認された。



(a)次元評定



(b)情動カテゴリー判断

図2. 顔のみ、映像のみを呈示されたときの(a)次元評定の平均値ならびに(b)各情動カテゴリーの選択率。

2.2 本実験

尺度評定および情動カテゴリー判断のいずれにおいても演者の意図した通りに情動が判断されたことを確認したため、収録された動画と音声の各4素材を、口の動きと音声同期するように総当たりで合成し(Premiere Pro CS6, Adobeによる)、顔表情と音声が一致、不一致な刺激を計16サンプル(4表情 × 4音声)作成した。これらの刺激を用いて、同種あるいは異種の情動を表現した顔と声を組み合わせるときに、観察者がいかなる情動を知覚するかを調べた。

2.2.1 実験参加者

36人の大学生(男性18人, 女性18人, 21~25歳, $M = 22.08, SD = 0.89$)が実験に参加した。いずれの参加者も刺激の妥当性確認実験には参加していなかった。

2.2.2 手続き

インフォームド・コンセントとして実験前に実験参加の同意を得た後、練習試行を行った。実験はブロックデザインで行い、音声はヘッドフォン(MM-HSUSB7BK, サンワサプライ)から呈示した。第1ブロックでは、(1)パソコンのディスプレイ上(VPCL23AJ, Sony)に5秒間表示される注視点(「+」)の中心にカーソルを持ってこること、(2)注視点の後、同一の刺激が連続で3回呈示されている間は、刺激を注視しマウスは動かさないこと、(3)刺激終了後ただちに表示される視覚的アナログスケール(「不快—快」の「活動性低—活動性高」どちらか一方)上で、演者が表現している情動の程度をマウスで左クリックすることを求めた。これを16の刺激それぞれについて繰り返した。16の刺激はランダムに呈示した。第2ブロックでは、第1ブロックと同様の手続きで、第1ブロックでは尋ねなかった方の尺度について回答を求めた。第1ブロックと第2ブロックの順序はカウンターバランスした。なお、「快・不快」および「活動性」の意味は、該当するブロックの前に口頭で説明した。第3ブロックでは、第1ブロックにおける手順(1)と手順(2)を行った後、右上方に「喜び」、左上方に「怒り」、左下方に「悲しみ」、右下方に「リラックス」と書いた画像を呈示し、演者が表現したと思う情動をクリックして選択するよう求めた。以上の刺激呈示と回答の収集はSuperLab 4.5(Cedrus)を用いて行った。

2.2.3 反応時間の補正

評定尺度課題において、「刺激呈示終了後、視覚的アナログスケール上でマウスをクリックするまでにかかる時間」は、「刺激を視聴して脳が情動を判断するまでの時間」と「情動を判断してからマウスを動かしてクリックするまでの時間」の総計であると考えられる。ここでの興味は、同種あるいは異種の情動の視聴覚刺激を組み合わせることによって、観察者が情動を判断するまでにかかる時間がいかに異なるかを明らかにすることであるため、マウスの動きによる時間ラグを差し引いた反応時間を算出する必要がある。

マウスの動きにかかる時間を測定するため、各参加者に、ディスプレイの中央に表示した注視点上にマウスを置いたまま待機してもらい、その後視覚的アナログスケールが表れた時点ですばやく実験者から指示された点（スケール上の「-100（左端）」、「0（中点）」、「100（右端）」の各点）をクリックするように求めた（「キャリブレーション課題」）。この作業を2回繰り返し、「中点からマウスを動かしてクリックするまでの時間」を計6点について得た。作業への慣れを考慮し、2回目の作業について、スケールの端（-100, 100）をクリックするまでの時間の平均値 t_{tip} と中点（0）をクリックするまでの時間 t_0 を算出した。

ある地点から距離 D の地点にある対象にまでかかる運動の時間 T は数式1によって表される (Fitts & Peterson, 1964)。これをフィッツの法則という。

$$T = a + b \log_2(1 + D) \tag{1}$$

ここで、各参加者について、先ほど算出した t_0 と t_{tip} から、係数 a (数式2) と係数 b (数式3) を求める。

$$a = t_0 \tag{2}$$

$$b = \frac{t_{tip} - t_0}{\log_2 101} \tag{3}$$

以上の係数 a 、係数 b および評定値の絶対値を数式1の a 、 b 、 D にそれぞれ代入し、「マウスを動かして直線をクリックするまでの時間 T (理論値)」を算出した。この T 値を SuperLab で測定した反応時間から引くことで、「刺激を見て脳が情動を判断するまでの時間」(以下、単に「反応時間」と呼ぶ) を算出した。

3. 結果

3.1 カテゴリー判断

顔と声の組み合わせ16刺激に対するカテゴリー判断について、各情動カテゴリーの選択率を求め

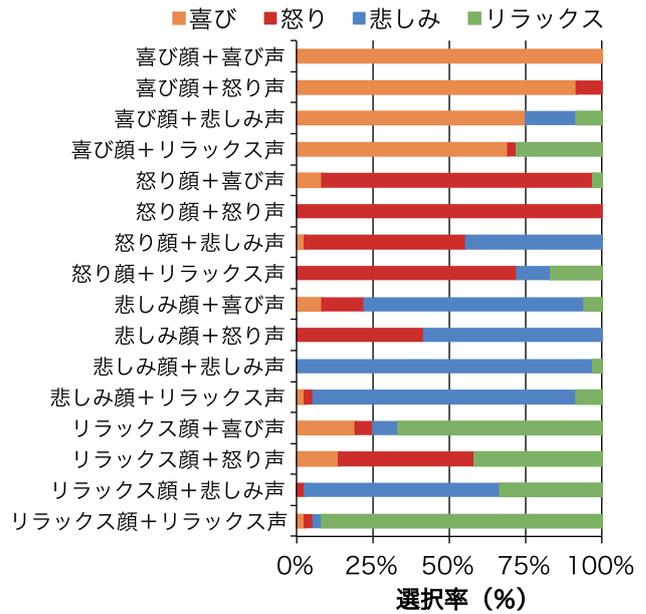


図3. カテゴリー判断における各情動カテゴリーの選択率。

た(図3)。図3から、「喜び顔」に対してはいずれの声を付加しても70%以上の参加者が「喜び」と判断したのに対し、「怒り顔+悲しみ声」、「悲しみ顔+怒り声」、「リラックス顔+怒り声」、あるいは「リラックス顔+悲しみ声」といった組み合わせでは、必ずしも顔で表現されている情動が判断されたわけではないようである。

これらの16種類の刺激に、刺激の妥当性確認実験における顔のみ4刺激、声のみ4刺激を加えた24種類の刺激に対して、いかなる情動が選択されたのかを明らかにするために単純コレスポネンス分析を行った。単純コレスポネンス分析は、二次元のクロス集計表を複数次元上の得点へと変換し、変数間あるいは変数内の相違や類似性を視覚的に表現する手法である (Greenacre, 1993)。ある特定の次元 α における情報量 (全体に対して次元 α が寄与する程度) は、次元 α に分解された χ^2 値の割合 (次元 α に対するイナーシャ λ_α^2 を全体に対するイナーシャ λ_T^2 で割った値) によって示される。全体に対するイナーシャ λ_T^2 はピアソンの適合度 χ^2 値をデータの観測数 ($N = 768$) で割った値に等しいため、ピアソンの χ^2 値から変数間の分布に偏りがあるかどうかを検定することができる (Greenacre, 1993)。

24の刺激と選択された情動の対応関係について2次元解を求め(図4, $\chi^2(69, N = 768) = 1537.22, p < .001$)、この2次元によってデータの70.40%が説明されたことがわかった(次元1: $\frac{\lambda_1^2}{\lambda_T^2} = .38$, 次元2: $\frac{\lambda_2^2}{\lambda_T^2} = .32$)。図4におい

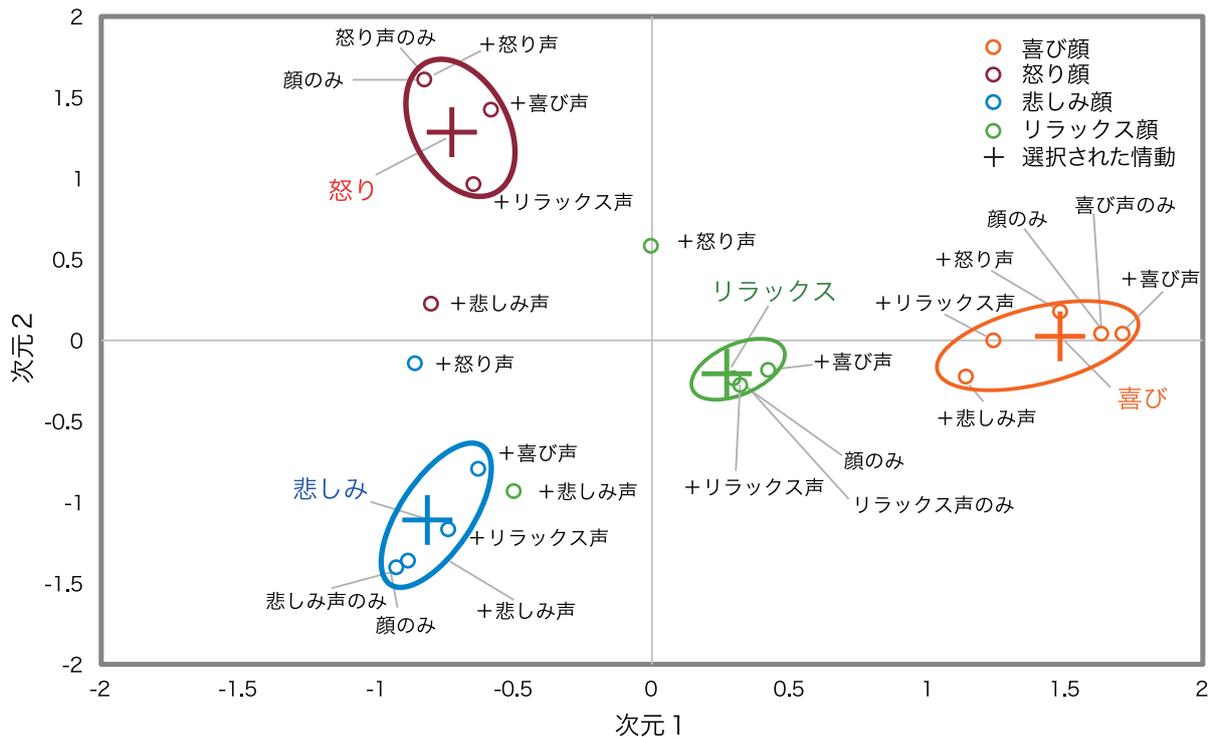


図4. 刺激（顔のみ4刺激, 声のみ4刺激, その組み合わせ16刺激の総計24刺激。で示す。）と実験参加者が選択した情動（喜び, 怒り, 悲しみ, 怒りのカテゴリー判断。+で示す。）の対応関係。各プロットは単純相関分析によって算出された次元得点を示す。組み合わせ16刺激の色分けは顔表情に基づいており、例えば青色の「+ 喜び声」と記されている場合は、悲しみの顔表情に喜びの声が付加された刺激を示す。

て、「+」で示された参加者の判断分布から、「喜び」が第1象限、「怒り」が第2象限、「悲しみ」が第3象限、「リラックス」が第4象限に分類され、Russell (1980) の2次元円環モデルと同様に、次元1が「快・不快」、次元2が「活動性」となるような配置が得られた。

図4上に、個々の情動カテゴリーについて、参加者の選択に近い刺激を楕円で示した。図2bおよび図3の通り、顔のみ、声のみ、あるいは顔と声で情動が一致している刺激については、参加者が選択した情動（「+」）と刺激（○）の位置がいずれも近いことが確認できる。

顔と声で情動が異なる刺激については、全体的に、顔で表される情動が選択される傾向が示された。「喜び顔」の場合は、いかなる情動の声が付加されても、「喜び」と選択される傾向があった。「怒り顔」、「悲しみ顔」についてもこの傾向が示されたが、「活動性」軸上で対となる情動の声が組み合わせられたとき（i.e., 怒り顔 + 悲しみ声, 悲しみ顔 + 怒り声）には声の表す情動を回答する参加者が増えることがわかった。この場合にも、「怒り顔 + 悲しみ声」は第2象限上に、「悲しみ顔 + 怒り声」は第3象限上に配置されているため、声より

も顔で示された情動が選択されることが多い傾向にあった。このような視覚優位の傾向は「リラックス顔」には認められず、リラックス顔に「快—不快」軸上で異種の性質を持つ声（i.e., 怒り声, 悲しみ声）が付加された場合には、声を表す情動を選択をする参加者が多かった。リラックス顔に「快—不快」軸上で同様の性質を持つ「喜び声」が付加された場合には「リラックス」と判断されたことがわかった。

以上の結果をまとめると、演者が表現した情動によって視聴覚相互作用の傾向が異なることが示された。喜び、怒り、悲しみについては視覚優位性が示され、いずれも顔で表された情動を観察者が判断する傾向があることがわかった。リラックスについては、不快な声が付加されたときに、声を表す情動を判断する傾向が示された。

3.2 次元評定

3.2.1 評定値

事前のデータの目視から、快刺激（喜び、リラックス）、不快刺激（怒り、悲しみ）のそれぞれが同

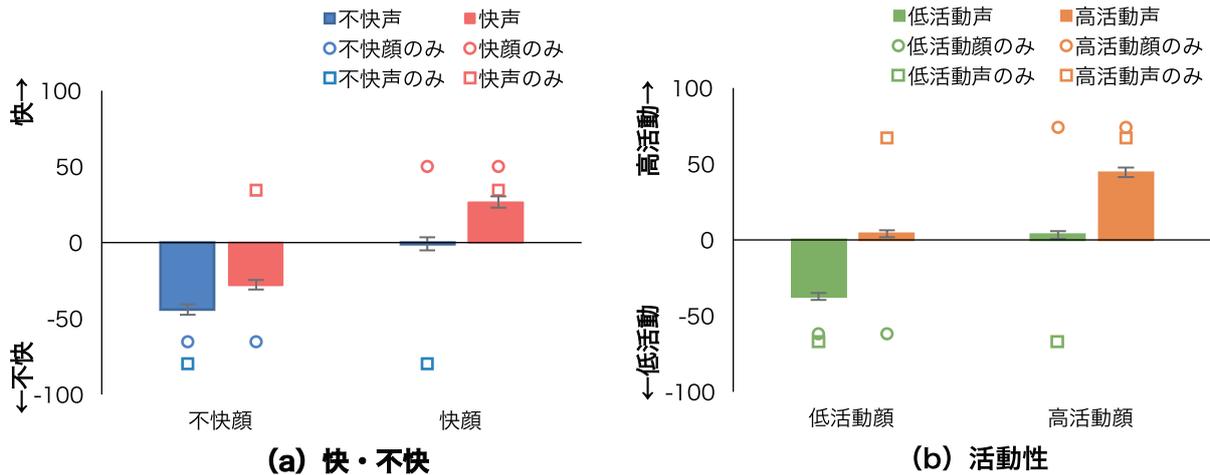


図5. 各刺激に対する(a)「快・不快」および(b)「活動性」の平均評定値。図中の各プロットは刺激の妥当性確認実験における顔のみ(○)および声のみ(□)に対する各評定の平均値を示す。エラーバーは標準誤差を示す。

様の傾向を示したため、「快・不快」評定値について、快刺激と不快刺激のそれぞれの平均値を算出した。同様に、高活動性刺激(喜び、怒り)と低活動性刺激(リラックス、悲しみ)のそれぞれも同様の傾向を示したため、「活動性」評定値について高活動性刺激と低活動性刺激のそれぞれの平均値を算出した(図5b)。以下、それぞれを快顔(快声)、不快顔(不快声)、高活動顔(高活動声)、低活動顔(低活動声)と呼ぶ。各条件において $M \pm 2SD$ から逸脱した評定があった場合にはその参加者を外れ値として分析からは取り除き、以下では、28人分のデータを分析した(図5)。各図中のプロットは、刺激の妥当性確認実験で得られた、顔のみ、声のみの平均評定値である。

「快・不快」次元(図5a)について、同種の情動が組み合わせられたときには(i.e., 不快顔+不快声、快顔+快声)、それぞれ演者が意図した通りの情動が評定されたが、異種の情動が組み合わせられたときには、それに比べて評定の絶対値が小さくなったようである。また、不快顔に快声が付加されたときには「不快」の方向に評定されているのに対し、快顔に不快音声が付加されたときには「ゼロ付近」に評定されており、モダリティによる非対称性が認められる。

これについて確認するため、顔と声を参加者内要因とする2要因分散分析を行ったところ、顔、声の主効果および二要因の交互作用がすべて有意であった(顔: $F(1, 27) = 71.97, p < .001, \eta_p^2 = .73$; 声: $F(1, 27) = 67.61, p < .001, \eta_p^2 = .72$; 顔×声: $F(1, 27) = 9.39, p = .005, \eta_p^2 = .26$)。Bonferroniの修正法を用いた多重比較の結果、不快顔および快顔におけ

る声の効果はいずれも有意であったが($ps < .001$)、前者の効果量($d = 0.92, 95\%CI [0.90, 0.94]$)の方が後者の効果量($d = 1.26, 95\%CI [1.24, 1.28]$)よりも小さかったことから、快顔よりも不快顔の方が付加する声の影響が小さいことがわかった。すなわち、不快な表情に快の音声を付け加えた場合には、その逆(快顔+不快声)に比べ、より「顔表情」で表出された情動の方向に判断されることがわかった。

「活動性」次元(図5b)については、同種の情動が組み合わせられたときには(i.e., 低活動顔+低活動声、高活動顔+高活動声)、演者が意図した通りの方向に評定され、異種の情動が組み合わせられたときには(i.e., 低活動顔+高活動声、高活動顔+低活動声)、中立(ゼロ)付近表情に評定されており、モダリティによる非対称性は認められなかった。「快・不快」情動と同様に、顔と声を参加者内要因とする2要因分散分析を行ったところ、顔および声の主効果はそれぞれ有意であったが、交互作用は有意でなかった(顔: $F(1, 27) = 203.40, p < .001, \eta_p^2 = .88$; 声: $F(1, 27) = 230.68, p < .001, \eta_p^2 = .90$; 顔×声: $F(1, 27) = 0.01, p = .94, \eta_p^2 < .001$)。各々の主効果の効果量がほぼ同じであったことから、受け手の情動評定に「表情」と「音声」が及ぼす影響が同程度であったことがわかった。

3.2.2 反応時間

フィッツの法則(Fitts & Peterson, 1964, 第2.2.3項を参照。)により補正した反応時間³について、

³この補正の際に行ったキャリブレーションよりも情動評定の方が判断が早かった場合があるため、この反応時間は真の値を取りうる。

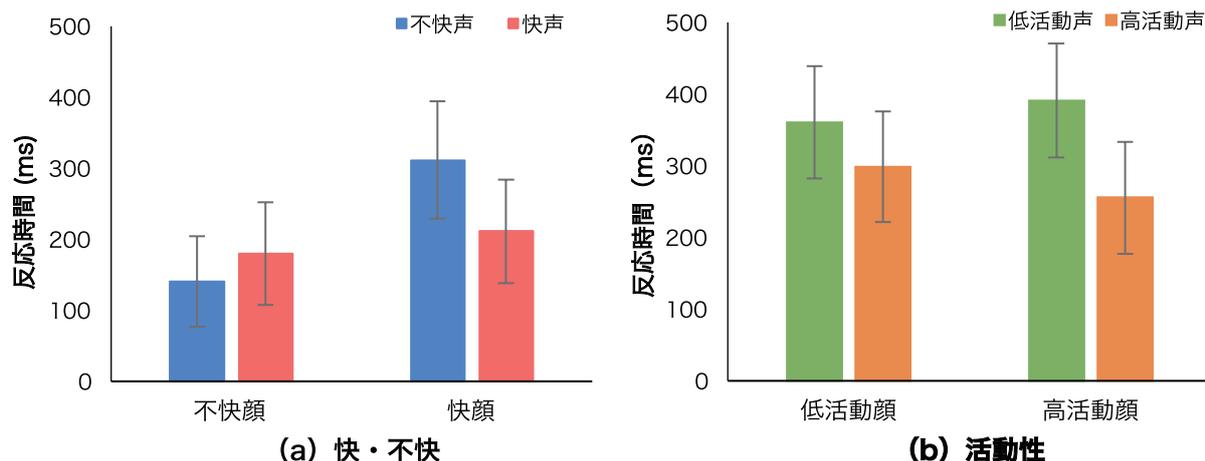


図6. 各刺激に対する(a)快・不快および(b)活動性の平均反応時間。エラーバーは標準誤差を示す。

各条件について $M \pm 2SD$ から逸脱した反応がなされた場合には、その参加者を外れ値として分析から除外した。各条件について分析対象とした27人分の反応時間の平均値を図6に示す。

図6aより、不快顔は快顔よりも反応が早かったが、声の情動によって反応時間に大きな違いはなかったようである。「快・不快」次元の評定にかかる反応時間について、 $2(\text{顔}) \times 2(\text{声})$ の2要因参加者内分散分析を行ったところ、顔の主効果のみが有意であり、声の主効果ならびに交互作用は有意でなかった(顔: $F(1, 26) = 5.67, p = .03, \eta_p^2 = .18$; 声: $F(1, 26) = 0.64, p = .43, \eta_p^2 = .02$; 顔 \times 声: $F(1, 26) = 2.00, p = .17, \eta_p^2 = .07$)。図6aから、快顔($M = 261.46, SD = 341.89$)よりも不快顔($M = 160.29, SD = 345.65$)の方が素早く反応されたことがわかった。

活動性(図6b)については、高活動顔に高活動声が付加されたときに低活動声を付加されたときよりも反応が早かったが、そうした声の効果は低活動顔では小さかったようである。活動性評定にかかる反応時間について、 $2(\text{顔}) \times 2(\text{声})$ の2要因参加者内分散分析を行ったところ、いずれの主効果も有意ではなかったが、交互作用は有意であった(顔: $F(1, 26) = 0.02, p = .89, \eta_p^2 = .001$; 声: $F(1, 26) = 0.58, p = .46, \eta_p^2 = .02$; 顔 \times 声: $F(1, 26) = 5.79, p = .02, \eta_p^2 = .18$)。Bonferroniの修正法を用いた多重比較により、低活動顔における声の効果は有意ではなかったが($p = .38$)、高活動顔における声の効果は有意であった($p = .03$)。図6bより、高活動顔について、低活動声よりも高活動声の評定にかかる反応時間の方が短いことがわかった。

3.2.3 評定値と反応時間の関係

各次元における参加者の情動の評定値と反応時間の関係を散布図として図7に示した。反応時間が $M \pm 2SD$ を逸脱した参加者は除外した。

刺激の種類(「不快と快」と「顔と声」の組み合わせ4種類ならびに「低活動と高活動」と「顔と声」の組み合わせ4種類)によって評定値と反応時間の関係が異なるのかをマルチレベル分析(階層線形モデルと同義。小野寺・岩田・菱村・長谷川・村山, 2006)によって調べた。具体的には、(1)刺激の種類によって、評定値から反応時間を説明する回帰直線の切片および回帰係数が異なるランダム切片・ランダム係数モデル、(2)刺激の種類によって回帰直線の切片が異なるランダム切片モデル、(3)刺激の種類による特定の関係がないというモデル(ヌルモデル)の3者のモデル比較を尺度ごと(i.e., 「快・不快」、「活動性」)に行った。その結果、「快・不快」については、ヌルモデルのデータのあてはまりが最もよかった(AICs = 1600.10(ランダム切片・ランダム係数モデル), 1596.10(ランダム切片モデル), 1594.70(ヌルモデル))が、回帰係数(標準化なし)は有意でなかった($B = -0.80, p = .47$)。「活動性」についてはランダム切片モデルが採択された(AICs = 1610.50(ランダム切片・ランダム係数モデル), 1607.00(ランダム切片モデル), 1608.20(ヌルモデル))。回帰係数を算出したところ $B = -3.22(p = .03)$ と、活動性が高ければ高いほど、より早く反応するという関係性が認められた。

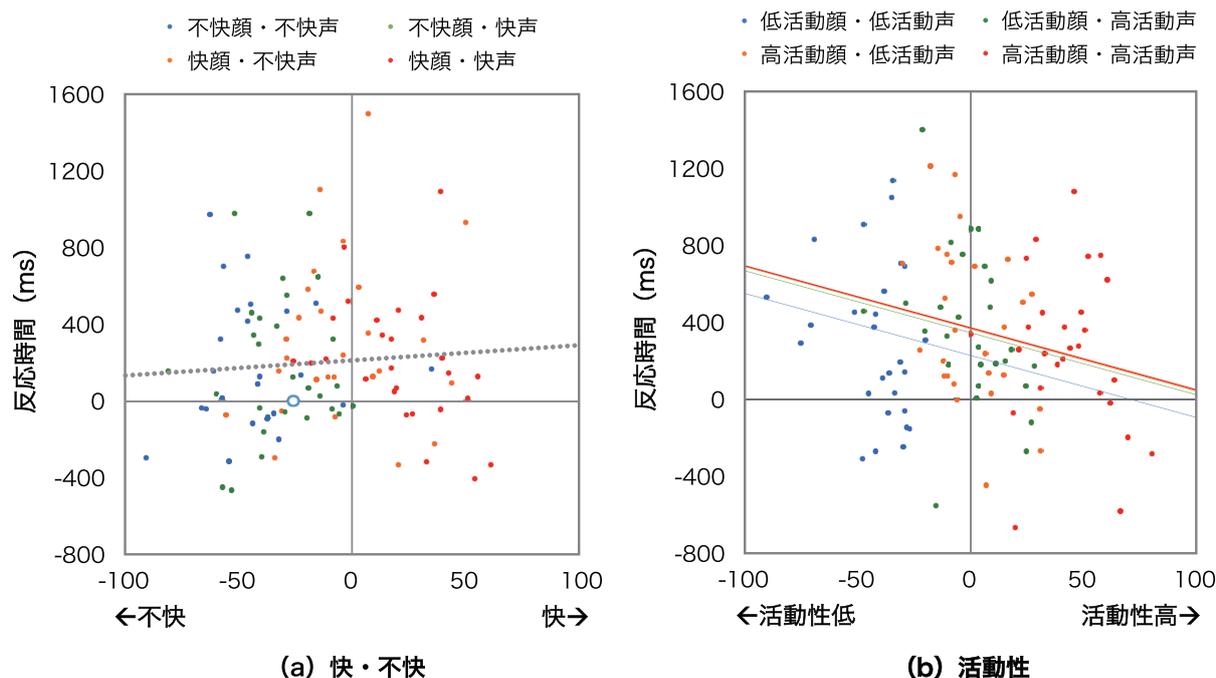


図7. 各刺激に対する評定値と反応時間の散布図((a)快・不快および(b)活動性の平均反応時間)。図中の直線は回帰直線を示す。(b)活動性ではランダム切片モデルが採択されたため(マルチレベル分析による)、刺激の種類ごとに回帰直線を引いている。

4. 考察

顔表情と音声で異なる情動を知覚した際の視聴覚相互作用について調べた。渡辺・望月(2004)は、情動判断における視覚優位性を示しているが、本研究はこれを概ね支持し、「リラックス」以外の情動については、視覚優位な傾向が認められた(情動カテゴリー判断による)。「リラックス」という情動がエクマンの基本情動には含まれていないように(e.g., Ekman & Friesen, 1975)、リラックスという顔表情が持つ情動の強度が十分でなかった可能性がある。また、不快表情である「怒り顔」と「悲しみ顔」については、活動性次元上で対となる「悲しみ声」と「怒り声」がそれぞれ付加されたときに、判断される情動が混同されやすいことが示された。ヒトが表情を知覚する際には「快・不快」を判断しやすいという知見に基づけば(高木ら, 2011)、副次的な情動である「活動性」については声の情動の影響を受けやすかったのだと考えられる。こうした影響が不快情動にのみ生じたのは、快表情の持つ「活動性」が情動カテゴリー判断では十分に区別されなかったことによる(図4)と考えられる。

尺度評定では、こうした視覚優位の傾向が「快・不快」次元には認められたが「活動性」次元には認められなかった。さらに不快表情の方が快表情よりも声の影響が小さいことを示した。ヒトは不

快な刺激に対してより素早く鋭敏に反応できるようプログラムされており(e.g., 戸田, 1987)、さらに視覚の方が聴覚よりも多くの情報を同時に処理できるため、不快表情の影響が大きかったのだと考えられる。こうした鋭敏な反応を要求されない「活動性」については、表情と音声伝える情動情報が同程度であったのだと考えられる。

反応時間については、表情と音声で異なる情動を組み合わせたときにも、「快・不快」では交互作用が認められず、「活動性」においても「低活動顔+低活動声」での反応時間が短くなったわけではないため、同種の刺激を組み合わせの方がより早く反応されるという渡辺・望月(2004)の知見は支持されなかった。「快・不快」については視覚的な不快刺激により素早く反応するという傾向が認められ、「活動性」に関しても、視聴覚の組み合わせは観察者の反応の早さには影響せず、むしろ「観察者が刺激をどの程度活動的だと思ったか」ということが反応時間を決定していることがわかった(図7b)。ヒトが情動を処理する際には、まず刺激を感覚として知覚し、その後、情動内容の総合的な判断を行う、という2つの過程が仮定されている(伊藤・梅本・山鳥・小野・往住・池田, 1994; 渡辺・望月, 2004)。不快表情は、前者の「感覚刺激として知覚する段階」で処理されるが(吉川・森崎, 1999)、その一方で、副次的な情動である「活

動性」はその後の“判断”の段階で処理されると考えられる。そのため、「快・不快」刺激に対しては観察者の“判断”と反応時間との間に明確な関係が認められなかったのに対し(図7a)、「活動性」刺激では認められたのだと考えられる。これらの結果は、「快・不快」と「活動性」という情報がそれぞれ独立に処理されていると考えたRussell(1980)の知見を支持している。

今後、他の話し手(女性の話し手や演劇経験のない一般の話し手)や話の内容によっても同様の結果が再現されるのかを確認する必要がある。また、我々が現実場面で経験する「明るい声だが目が笑っていない」というような話し手の表現の特徴を探索することで、より生態学的に妥当な状況における話し手から受け手のコミュニケーションの全貌を明らかにすることができる。さらには、音楽演奏のような、特定の論理的・言語的意味を持たない素材を対象とすることで、ヒトの情動情報処理における視覚情報と聴覚情報の役割をそれぞれ明らかにすることができるであろう。

謝辞

刺激作成にご協力いただいた演者に感謝いたします。また本研究に対して貴重なご助言をいただいた田部井賢一氏(三重大学)に感謝いたします。

本研究は第4著者が同志社大学文化情報学部に提出した卒業論文(「感情認知におけるクロスモーダル性に関する検討」)に基づき、追加実験と再分析を行ったものである。

参考文献

- [1] 荒井宏太・井上康之・小野和也・板倉昭二・北崎充晃(2011). 表情と無意味語音声のクロスモーダル情動認知: モダリティ情報の強度と信頼性の効果の検討. *認知科学*, 18(3), 428-440.
- [2] 大坊郁夫(1998). *しぐさのコミュニケーション: 人は親しみをどう伝えあうか*. サイエンス社.
- [3] Ekman, P., & Friesen, W. V. (1975). *Unmasking the face: A guide to recognizing emotions from facial expressions*. Englewood Cliffs, NJ: Prentice Hall.
- [4] Fitts, P. M., & Peterson, J. R. (1964). Information capacity of discrete motor responses. *Journal of Experimental Psychology*, 67(2), 103-112. doi: 10.1037/h0045689.
- [5] Greenacre, M. J. (1993). *Correspondence analysis in practice*. London: Academic Press.
- [6] Mehrabian, A. (1972). *Nonverbal communication*. Chicago, IL, Aldine-Atherton.
- [7] Ohgushi, K., & Hattori, M. (1996). Emotional communication in performance of vocal music. In B. Penneycook, & E. Costa-Giomi (Eds.), *Proceedings of the Fourth International Conference on Music Perception and Cognition* (pp. 269-274). Montreal: McGill University.
- [8] 小野寺孝義・岩田昇・菱村豊・長谷川孝治・村山航(2006). *基礎から学ぶマルチレベルモデル*. ナカニシヤ出版.
- [9] Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality & Social Psychology*, 39(6), 1161-1178. doi: 10.1037/h0077714.
- [10] 田中章浩(2011). 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*, 18(3), 416-427.
- [11] Tanaka, A., Koizumi, A., Imai, H., Hiramatsu, S., Hiramoto, E., & de Gelder, B. (2010). I feel your voice. Cultural differences in the multisensory perception of emotion. *Psychological Science*, 1259-1262.
- [12] 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. *電子情報通信学会技術研究報告*, 111(283), 51-56.
- [13] 戸田正直(1987). 心をもった機械: ソフトウェアとしての「感情」システム. *ダイヤモンド社*.
- [14] 渡辺桃子・望月登志子(2004). 表情認知における視聴覚情報の相互規定性. *感情心理学研究*, 11(2), 53-64.
- [15] 吉川左紀子・森崎礼子(1999). 表情による感情メッセージの伝達. *電子情報通信学会技術研究報告*, 98(503), 31-38.