

情動認知におけるクロスモーダル性に関する検討

岸本和香[†] 正田悠^{†,‡} 新田晴[†] 阪田真己子[†]
同志社大学[†] 日本学術振興会特別研究員[‡]

1. はじめに

渡辺ら (2004) は、話し手が顔表情と音声で異なる情動を表出したとき、受け手は表情に基づいて情動を判断し (e.g., 悲しみの表情に喜びの声が付随するとき、聴き手は「悲しみ」の情動を知覚する)、表情と音声が一歩一致している場合よりも処理に時間がかかることを示している。先行研究では Ekman の基本 6 情動を用いているため、個々の情動 (以下、情動カテゴリー) のいかなる特徴がこうした視聴覚統合に影響しているかが十分に明らかでない。一般にヒトが表出・知覚する情動は 2 次元 (以下、「快-不快」「活動性」) で表すことができ、この 2 次元はそれぞれ独立なものであると考えられている (Russell, 1980)。本研究では、情動 2 次元上に布置する 4 つの情動カテゴリー (リラックス、喜び、怒り、悲しみ) を顔表情と音声で表出してもらい、これらの情報を組み合わせたときに受け手がいかなる情動を受け取るかをカテゴリー判断ならびに次元評定から調べた。音声よりも表情の方が快-不快を判断しやすく (高木ら, 2011)、ネガティブな情報に対して敏感 (渡辺ら, 2004) な傾向があることから、不快な表情を見ている際には音声の影響は弱いと予測される。また先行研究と一致して、異なる情動が視聴覚統合されたときには、同一の情動が統合されたときよりも処理に時間がかかると予測される。

2. 方法

2.1. 参加者

36 名の大学生 (男女各 18 名, 21~25 歳, $M = 22.08$, $SD = 0.89$) が実験に参加した。

2.2. 刺激

演劇活動経験 8 年以上の男性 (28 歳) および女性 (29 歳) 各一名に無意味音節「にれなこ」を 4 つの情動カテゴリー (上述) で表現してもらった。大学生 10 名に動画および音声をそれぞれ呈示し、意図した情動が表現されていることを確認した (正答率 7 割以上)。動画と音声を合成し、各演者について表情と音声が一歩一致・不一致な刺激を計 16 (4 表情×4 音声) 作成した。

2.3. 手続き

情動カテゴリー (i.e., リラックス、喜び、怒り、

悲しみ) の強制選択課題、「快-不快」「活動性高-活動性低」の 2 項目の評定課題 (-100~100 の視覚的アナログスケール) からなる実験を行った。刺激はディスプレイ上で呈示し、刺激終了後表示される情動カテゴリー (強制選択) または数直線 (評定尺度) をマウスでクリックさせることで回答を収集した。刺激呈示終了後から回答までの時間を計測し、マウスの動きによる時間ラグを考慮した補正值を分析に用いた。

3. 結果

3.1. 情動カテゴリー判断

表情と音声と同じ情動を表現している刺激については、表現された情動と一致した情動が判断された (全体の 96.18%)。一方、表情と音声で情動が異なる刺激については、表情に規定されやすい情動と音声に規定されやすい情動があった (表情が表す情動が 67.13%, 音声を表す情動が 24.42%, どちらもでない情動が 8.45% 選択された)。そこで、表情と音声の情動が異なる刺激について、参加者が表情と音声のどちらの情動を回答したのかを集計した (表 1)。表情刺激について χ^2 検定を行った結果、度数に有意な偏りが認められた ($\chi^2 (3, N = 791) = 72.86$, $p < .001$)。残差分析の結果、表情が「リラックス」のとき音声の情動を、「喜び」および「悲しみ」のときに表情の情動を回答する割合が有意に高かった (いずれも $p < .001$)。

また、音声刺激について χ^2 検定を行った結果、度数に有意な偏りが認められた ($\chi^2 (3, N = 791) = 83.99$, $p < .001$)。残差分析の結果、音声で「リラックス」および「喜び」のとき表情の情動を、音声で「悲しみ」のときに音声の情動を回答する割合が有意に高かった (いずれも $p < .001$)。

表 1. 各刺激のクロス集計表

		回答		合計
		音声	表情	
表情刺激	リラックス	96	97	193
	喜び	35	166	201
	怒り	50	145	195
	悲しみ	30	172	202
合計		211	580	791
音声刺激	リラックス	28	165	193
	喜び	33	165	198
	怒り	48	150	198
	悲しみ	102	100	202
合計		211	580	791

Exploring the cross-modality in the perception of emotion
Waka Kishimoto[†], Haruka Shoda^{†,‡}, Haru Nitta[†], Mamiko Sakata[†]
[†]Doshisha University, [‡]JSPS Research Fellow

以上から、表情で表出された情動が判断される傾向が示された。その中で「リラックス」表情は音声の影響を受けやすく、「悲しみ」音声は表情に影響されにくいことが示された。

3.2. 情動次元評定

3.2.1. 評定値

「快-不快」評定値について快刺激（リラックス、喜び）および不快刺激（悲しみ、怒り）の平均値を、「活動性」評定値について高活動性刺激（喜び、怒り）および低活動性刺激（リラックス、悲しみ）の平均値を算出し、尺度（快-不快、活動性）ごとに 2（表情）×2（音声）の一般化多変量分散分析（GMANOVA）を行った。

「快-不快」次元（図 1a）については表情、音声の主効果および交互作用がすべて有意であった（表情：Wilks' $\lambda = .33, F(1, 35) = 70.44, p < .001, \eta_p^2 = .67$ ；音声：Wilks' $\lambda = .36, F(1, 35) = 62.09, p < .001, \eta_p^2 = .64$ ；表情×音声：Wilks' $\lambda = .78, F(1, 35) = 9.69, p = .001, \eta_p^2 = .22$ ）。Bonferroni の修正法を用いた多重比較の結果、快表情と不快音声を組み合わせると、快-不快の評定が中立（ゼロ付近）になり、不快表情と快音声を組み合わせると、不快方向の評定値が高かった。これは、表情、音声に関わらず「不快」情報の影響力が大きいことを示している。

「活動性」次元（図 1b）においては、表情、音声の主効果は有意であったが、交互作用は有意でなかった（表情：Wilks' $\lambda = .16, F(1, 35) = 189.83, p < .001, \eta_p^2 = .84$ ；音声：Wilks' $\lambda = .11, F(1, 35) = 274.95, p < .001, \eta_p^2 = .89$ ）。また、効果量（ η_p^2 ）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。これに起因して、異なる情動の表情と音声を組み合わせたときに評定値が中立の値になったのだと考えられる（図 1b）。

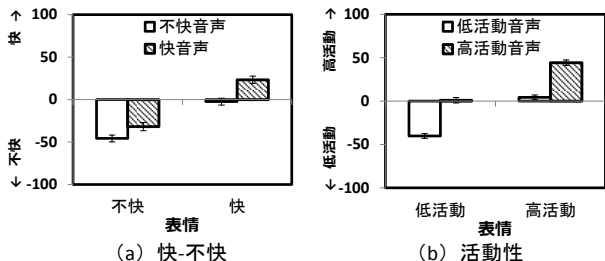


図 1. 情動次元の評定値

3.2.2. 反応時間

2（次元）×2（表情）×2（音声）の GMANOVA の結果、次元の主効果が有意傾向（Wilks' $\lambda = .92, F(1, 35) = 2.97, p = .09, \eta_p^2 = .08$ ）、表情と音声の二要因交互作用が有意で

あった（Wilks' $\lambda = .71, F(1, 35) = 14.19, p = .001, \eta_p^2 = .29$ ）。快-不快の判断よりも活動性の判断の方に時間がかかり（図 2）、次元上で逆方向の情動が組み合わせると判断に時間がかかる（図 3）ことがわかった。

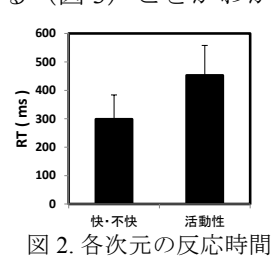


図 2. 各次元の反応時間

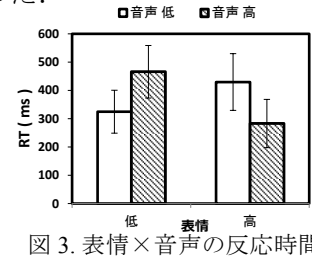


図 3. 表情×音声の反応時間

4. 考察

顔表情と音声で異なる情動を知覚した際のクロスモーダルな相互作用について調べた。渡辺ら（2004）は、情動判断における視覚優位性を示しているが、本研究はこれが「快-不快」次元にのみ認められ「活動性」次元には認められないことを示した。さらに「快-不快」次元における音声の影響が非対称性を示すことも明らかにした（「顔は笑っているが声は冷たい」ときにもポジティブには判断されない）。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いるため（快-不快次元の反応時間が短いことから確認できる）、聴覚よりも多くの情報を同時に処理できる視覚の方が優位に働くのだと考えられる。こうした素早い反応を要求されない「活動性」については、表情と音声で伝える情動情報が同程度であったのだと考えられる。

また、表情と音声で異なる情動を組み合わせたときには「快-不快」か「活動性」かに関わらず認知的コンフリクトが生じ、情動判断により時間がかかることを示した。個別の情動に対する事前分析により、両次元において相対する情動が組み合わせられた場合には（たとえば悲しみと喜び）、いずれかの次元が一致している場合よりも反応時間が遅くなることが示されている。これは、受け手が複数モダリティ（表情、音声）における情動情報を加算的に処理していることを示唆している。今後、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」というような状況において、話し手の表現と受け手の判断の特徴を探索することで、情動情報の視聴覚統合の全貌を明らかにすることができると考えられる。

参考文献

- 高木他 (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.
- Russell, J. A. (1980). A circumplex model of affect. *J. of Pers. Soc. Psychol.*, 39(6), 1161-1178.
- 渡辺他 (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.

謝辞：刺激作成にご協力いただいた演者 2 名ならびに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。