

# 発話と映像的身振りの統合的理解における聞き手の視線

## The Listener's Gaze in the Process of Gesture-Speech Integration

三宅 英典<sup>†</sup>, 関根 和生<sup>‡</sup>  
Hidenori Miyake, Kazuki Sekine

<sup>†</sup>松山東雲女子大学, <sup>‡</sup>慶應義塾大学  
Matsuyama Shinonome College, Keio University  
h.forte.rc@gmail.com

### 概要

本研究は、発話と身振りの処理過程を明らかにするために、聞き手に対する視覚的注意に焦点を当て、子どもが話者のどこを見ているかを検討した。6歳児21名、女子大学生21名を対象に、日常的な動作を発話と身振りで伝達するビデオをみせ、その後、4枚の写真からメッセージと最も一致するものを選択させた。課題に取り組む参加者の視線を計測した結果、6歳児は顔や身振りに対する視線に成人と差がみられなかったものの、発話と身振りの統合能力では両者に差があることが明らかになった。

キーワード：発話、身振り、統合的理解、視線、幼児期

### 1. 問題と目的

コミュニケーション場面において、聞き手は発話だけでなく身振りも考慮してメッセージを理解している[1]。しばしば、身振りは、発話では表されない情報を伝達する。例えば、数あるうちのリングから1つのリングを指さしながら「大きいね」と言う場合（指さしが対象物を特定）、や「走って逃げちゃった」と言いながら腕を振って人物の走る様子を描写する場合などである（身振りが人物の走り方を特定）。本研究では、このように発話と身振りがそれぞれ独自の情報を含む際、

その両方の情報を統合して行うメッセージ理解を統合的理解と呼ぶ。

例えば、Sekine et al. [2] は、3歳児・5歳児・成人を対象に、日常的な動作を表すメッセージについて発話と身振りで提示し（図1のように「乗っています」と言いながら両手でハンドルを掴む身振りをする）、その後、メッセージと最も一致する選択肢を4枚の写真から選ばせた。結果として、5歳児は、発話と身振りを成人と同程度に統合できるが、3歳児は両者の統合が難しいことが明らかになった。Miyake & Sugimura [3] では、同じような場面で身振りを参照する指示語発話（例：「こうやって」）を付加すると、幼児の統合的理解が促進されることを明らかにした。

このように先行研究では、幼児期の間に子どもは発話と身振りの情報を統合して話者のメッセージを理解することができるようになる、ということがわかってきた。だが、幼児は実際にどのように発話と身振りの情報を処理しているか、その処理過程に関しては不明な点が残されている。そこで、本研究は、発話と身振りの処理過程を明らかにするために、聞き手に対する視覚的注意に焦点を当て、子どもが話者のどこを見ているのかということを検討した。

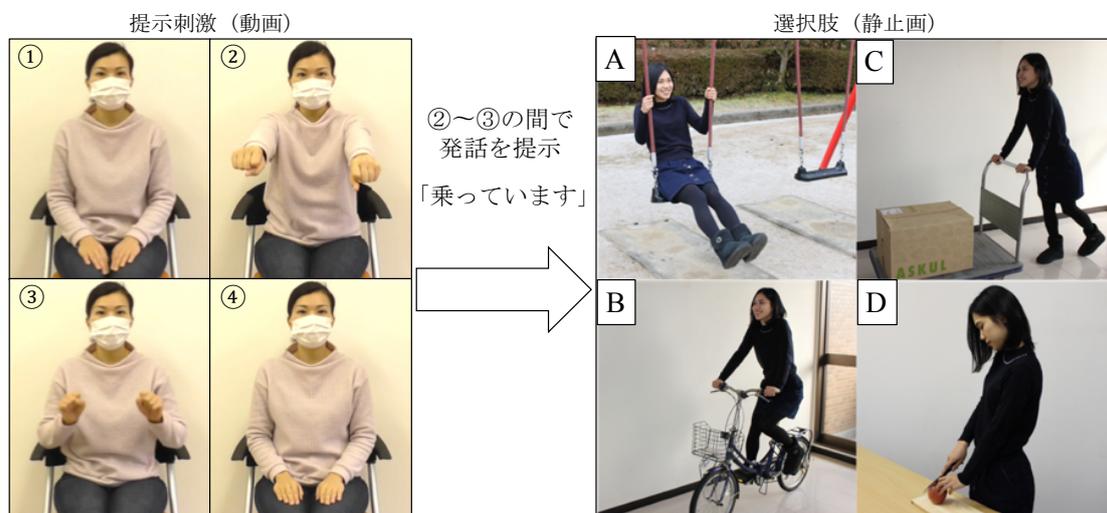


図1 発話と身振りの提示方法と選択肢の写真の一例

## 2. 方法

参加者は、6歳児21名（女児9名、平均6歳2ヶ月）と女子大学生21名（平均19歳）であった。課題は、モデルが日常的な動作を発話と身振りで伝達するビデオをみて、その後、4枚の写真からメッセージと最も一致するものを選択するものであった（図1）。課題で使用した動作は、先行研究にならい [2] [3]、書く・投げる・乗る・読む・飲む・開ける・食べる・登るの8つを用いた。これらの動作は、日本語版マッカーサー乳幼児言語発達質問紙によって3歳児の80%以上が理解している [4] と判断された動詞をもとに選出された。提示刺激は次の3条件で参加者に提示した（図1）。モデルが身振りと言語の両方を提示する発話-身振り条件（VG条件）、モデルの静止画（図1の①）と言語を提示する発話条件（V条件）、発話をミュートにしてモデルの動画のみを提示する身振り条件（G条件）。課題の反応をみるための4枚の写真（選択肢）は、モデルが発したメッセージに対して発話情報のみと一致する発話選択肢（図1のA）、身振り情報のみと一致する身振り選択肢（B）、発話と身振りの情報と一致する統合選択肢（C）、発話と身振りのどちらの情報とも一致しない無関係選択肢（D）であった。

課題の試行数は、練習課題で6試行（2動作×3条件）、本試行は18試行（6動作×3条件）であった。課題の提示順について、動作の提示順は固定し、3つの提示条件と選択肢の位置はカウンターバランスをとった。参加者の視線を計測するためにトビー・テクノロジー社製 Tobii nano を使用した。刺激は14インチのラップトップで提示し、参加者の課題遂行中における視線を計測した。分析ソフトには同社製 Tobii Pro Lab を使用した。

発話と身振りの統合を示す指標として、本研究は Multi Modal Gain Score (MMG 値) を測定する [2] [3] [5] [6]。MMG 値は以下のようにして算出する。

MMG 値 = VG 条件における統合選択肢の選択割合 - (V 条件と G 条件で、統合選択肢の選択割合が高い方の値)

聞き手が発話と身振りを統合しているのであれば、統合選択肢の選択割合は、VG 条件が他の条件と比べて最も高くなるはずである。そのため、聞き手が発話と身振りの統合的理解に基づいて課題に回答した場合、

MMG の値は1に近づくが、そうでない場合にはMMG は0に近い値をとる。

## 3. 結果

**課題の正答率と MMG 値** 参加者の年齢群ごとに課題の正答率と MMG 値を算出した（表1）。各条件における正答の選択肢は V 条件が発話選択肢と統合選択肢、G 条件が身振り選択肢と統合選択肢、VG 条件が統合選択肢であった。

表1 各条件の正答率とMMG値の平均値 (SD)

	6歳児	成人
V	0.79 (0.20)	1.00 (0)
G	0.87 (0.14)	0.94 (0.10)
VG	0.65 (0.22)	0.94 (0.11)
MMG	0.10 (0.24)	0.19 (0.17)

各条件の正答率に対して、チャンスレベルの検定を行った（V 条件、G 条件は50%、VG 条件は25%に設定した）。その結果、各条件の正答率は、6歳児と成人においてチャンスレベルを1%水準で有意に上回った。次に、MMG 値が0を超えているかを判断するためにチャンスレベルの検定を行った。その結果、6歳児はチャンスレベルを有意傾向で上回り ( $t(20) = 1.98, p < .10$ )、成人はチャンスレベルを1%水準で有意に上回っていた。

VG 条件の正答率に対して、年齢群（6歳児 vs. 成人）の  $t$  検定を行ったところ、群間に有意な差がみられた ( $t(40) = 5.41, p < .01$ )。具体的には、課題の正答率は成人の方が6歳児よりも高かった。一方、MMG 値に対して、年齢群（6歳児 vs. 成人）の  $t$  検定を行ったところ、群間に有意差はみられなかった。

**課題遂行時における参加者の視線の分析** 参加者の視線を計測するために、本試行で提示したすべての動画刺激において、Area of Interest (AOI) を顔領域と身振り領域に設定した（図2）。AOI の注視時間 (Duration) と AOI 内における注視点 (Fixation) を計測した。注視時間の定義は AOI 内の合計注視時間（ミリ秒）とし、注視点の定義は視線の停留時間が200ミリ秒で1点とした。

年齢群別に条件ごとの注視点と注視時間の平均値を

算出した (表 2). 提示刺激に対する聞き手の視線を検討するために, VG 条件の注視点と注視時間を従属変数として, それぞれに対して, 年齢群 (6 歳児, 成人)



図 2 「食べる」動作における顔と身振りの Area of Interest (AOI)

× AOI (顔, 身振り) の 2 要因分散分析を行った. その結果, 注視点と注視時間で, AOI の主効果がみられた (順に,  $F(1, 40) = 8.15, p = .01, partial \eta^2 = .17$ ;  $F(1, 40) = 14.27, p = .01, partial \eta^2 = .26$ ). 具体的に, 身振りよりも顔に対する注視時間が長く, 注視点の数も多かった. 6 歳児と成人の間に, AOI に対する注視点や注視時間の差はみられなかった.

VG 条件における視線と MMG 値の相関分析 6 歳児と成人に分けて, VG 条件における視線データと MMG 値の間で相関分析を行った (表 3,4). 6 歳児では, 顔の注視点と顔の注視時間, 身振りの注視点と身振りの注視時間に正の相関が示された. MMG 値と視線の間に有意な相関は示されなかった.

次に, 成人では, 顔の注視点と顔の注視時間, 身振りの注視点と身振りの注視時間に正の相関が示された. MMG 値は, 顔の注視点との間で正の相関を有意傾向で示した (表 4).

課題で使用した動作項目の分析 本試行で使用した 6 つの動作項目 (乗る, 読む, 飲む, 開ける, 食べる, 登る) に分けて, VG 条件における注視点と注視時間の平均値を AOI 別に算出した (図 3,4,5,6). まず, 顔の注視点と注視時間の平均値について, 対象 (6 歳児, 成人) × 動作項目 (6 つ) の 2 要因分散分析を行った. その結果, 注視点と注視時間の両方で, 動作項目のみ主効果が有意であった (順に,  $F(5, 200) = 8.26, p = .01, partial \eta^2 = .17$ ;  $F(5, 200) = 5.54, p = .01, partial \eta^2 = .12$ ).

動作項目の主効果に対して Holm 法による多重比較を行ったところ, 顔の注視点では, 「乗る」が「開ける」より, 「飲む」が「乗る, 読む, 開ける」より, 「食べる」が「開ける」よりも多かった (順に,  $t(40) = 3.99, p < .01$ ;  $t(40) = 3.46, p < .05$ ;  $t(40) = 6.17, p < .01$ ;  $t(40) = 4.18, p < .01$ ;  $t(40) = 4.24, p < .01$ ). 顔の注視時間で

表2 条件に分けたAOIに対する注視点の割合と注視時間の平均値 (SD)

条件	視線	AOI	6歳児	成人
VG	注視点	顔	0.44 (0.17)	0.40 (0.25)
		身振り	0.31 (0.26)	0.26 (0.14)
	注視時間	顔	1605.05 (784.05)	1667.86 (1089.67)
		身振り	981.03 (473.92)	1062.25 (597.03)
V	注視点	顔	0.56 (0.20)	0.46 (0.30)
		身振り	0.07 (0.06)	0.06 (0.06)
	注視時間	顔	1921.96 (953.38)	1959.60 (1307.16)
		身振り	220.52 (188.24)	236.51 (224.47)
G	注視点	顔	0.49 (0.17)	0.40 (0.25)
		身振り	0.28 (0.14)	0.30 (0.14)
	注視時間	顔	1782.08 (825.28)	1588.67 (1002.18)
		身振り	1023.58 (574.59)	1158.85 (598.94)

は、「飲む」が「読む、開ける、登る」よりも長かった (順に,  $t(40) = 3.76, p < .01$ ;  $t(40) = 4.90, p < .01$ ;  $t(40) = 3.11, p < .05$ ).

次に同様の分析を VG 条件の身振りに対する注視点と注視時間の平均値に対して行った. その結果, 注視点では対象と動作項目ともに主効果や交互作用が有意でなかった. 一方, 注視時間では, 動作項目の主効果と対象×動作項目の交互作用が有意であった (順に,  $F(5, 200) = 5.66, p = .01, partial \eta^2 = .12$ ;  $F(5, 200) = 2.66, p = .05, partial \eta^2 = .06$ ).

動作項目の主効果に対して Holm 法による多重比較を行ったところ, 身振りの注視時間では, 「読む」が「食べる、登る」よりも長かった (順に,  $t(40) = 3.96, p < .01$ ;  $t(40) = 4.21, p < .01$ ). 次に, 対象×動作項目の交互作用について, 6 歳児の注視時間は「乗る、読む、開ける」が「食べる」より長かった (順に,  $t(40) = 3.10, p < .05$ ;  $t(40) = 3.82, p < .01$ ;  $t(40) = 3.99, p < .01$ ). これ

に対して, 成人の注視時間は, 「読む」が「登る」より長かった ( $t(40) = 3.55, p < .05$ ). また, 「食べる」の身振りの注視時間のみ, 6 歳児よりも成人の方が長かった ( $F(1, 240) = 5.18, p = .05, partial \eta^2 = .11$ ).

表3 6歳児における各AOIの注視点・注視時間と課題の正答率の相関

		注視点		注視時間	
		顔	身振り	顔	身振り
注視点	顔				
	身振り				
注視時間	顔	-.086			
	身振り	.848 **	.023		
MMG	顔	-.062	.683 **	.199	
	身振り	-.135	-.055	-.153	-.165

\*\*  $p < .01$ , \*  $p < .05$ , +  $p < .10$

表4 成人における各AOIの注視点・注視時間と課題の正答率の相関

		注視点		注視時間	
		顔	身振り	顔	身振り
注視点	顔				
	身振り				
注視時間	顔	-.042			
	身振り	.987 **	-.061		
MMG	顔	.005	.975 **	.007	
	身振り	.375 +	-.167	.343	-.187

\*\*  $p < .01$ , \*  $p < .05$ , +  $p < .10$

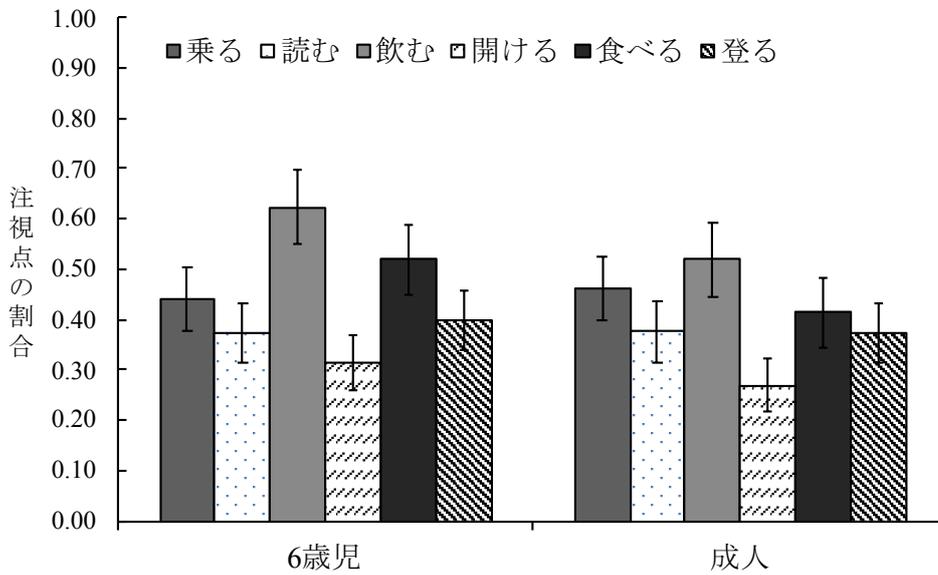


図3 身振り一顔 AOI 内のうち顔に対する注視点を動作項目別に分けた割合の平均値

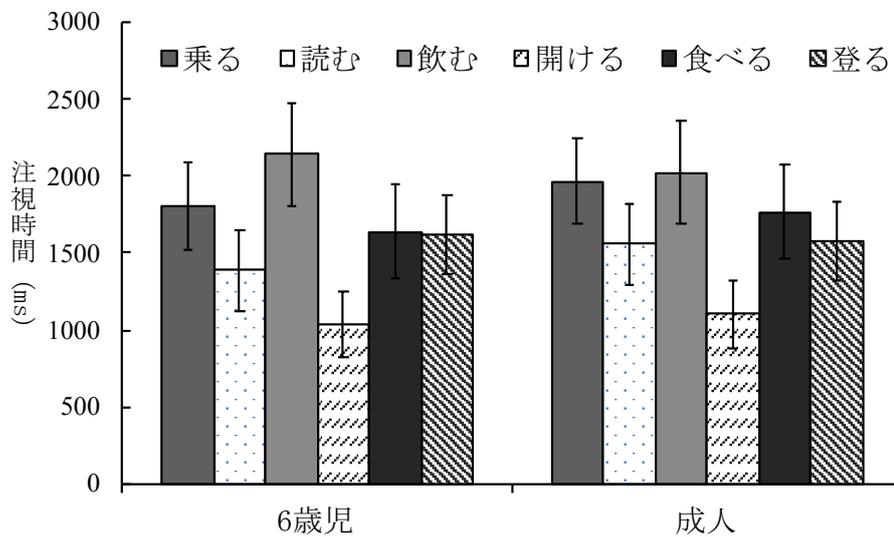


図4 顔 AOI 内における注視時間を動作項目別に分けた割合の平均値

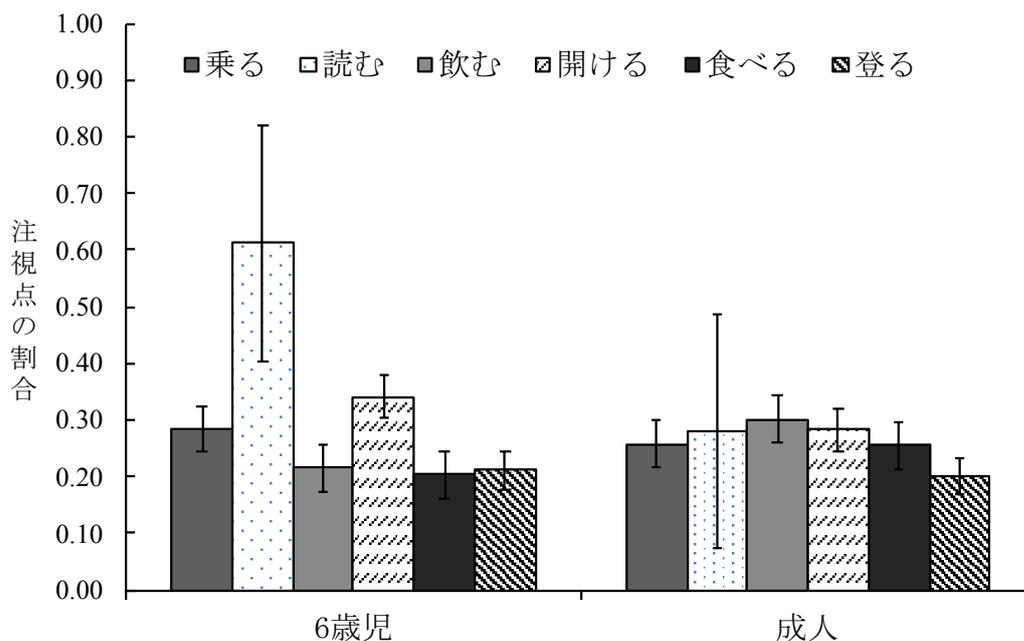


図5 身振り一顔 AOI 内のうち身振りに対する注視点を動作項目別に分けた割合の平均値

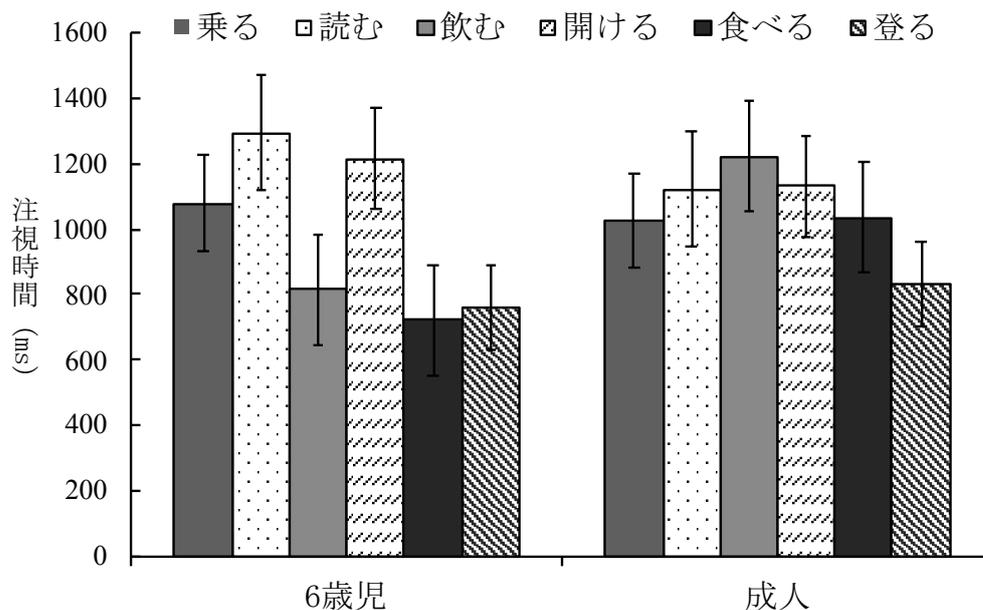


図6 身振り AOI における注視時間を動作項目別に分けた割合の平均値

#### 4. 考察

**課題の正答率** MMG 値が 0 を有意に超えていたのは成人のみで、6 歳児は有意傾向であった。しかしながら、*t* 検定の結果では、成人と 6 歳児の間で MMG 値に有意差がみられなかった。また、VG 条件の正答率は

成人の方が 6 歳児よりも高かった。*t* 検定だけでなく、個別に MMG 値を分析した場合や VG 条件の正答率を考慮すると、本研究における 6 歳児の身振り一発話統合能力は成人と同程度ではないものの、その能力の発達過程にあると考えられる。

**発話と身振りを見る聞き手の視線** 聞き手は、話者のメッセージを理解する際に身振りよりも顔に対する注視点が多く、注視時間も長かった。この傾向は 6 歳児

と成人の両方で同程度にみられた。このことから、聞き手が発話と身振りの統合的理解を行う際、身振りよりも顔に注意を向けながら話者のメッセージを理解しようとしていることが明らかになった。

**各条件の視線とMMG値の相関分析** 6歳児は、MMG値と視線の間に有意な相関がみられなかった。一方、成人では、MMG値が顔の注視点との間に有意傾向で正の相関を示した。そのため、発話と身振りを統合できるようにするためには、話者の顔に対する視覚的注意を高めることが重要であるかもしれない。

ただし、顔や身振りに対する注視点の数や注視時間は、年齢群（6歳児、成人）に主効果がみられなかった。つまり本研究の結果は、6歳児と成人の両者が、同程度に話者に対する視覚的注意を向けていたにもかかわらず、発話と身振りの統合能力との関連が現れたのは成人のみであったことを示している。これは、発話と身振りの処理過程において、6歳児が成人と同様の注意の向け方をしていることを示しているとともに、発話と身振りの情報を認知的に処理する過程では両者の間に差があることを示唆している。

**動作項目別に分けた視線の分析** 参加者の視覚的注意を動作項目別に分析したところ、動作項目に主効果がみられ、動作項目によって顔や身振りに対する注意に差があることが明らかになった。例えば、顔の注視点では、「乗る」が「開ける」より、「飲む」が「乗る、読む、開ける」より、「食べる」が「開ける」よりも多かった。顔の注視時間では「飲む」が「読む、開ける、登る」よりも長かった。「飲む」や「食べる」は、それぞれの身振りが顔に向かっていく動作を含んでいる。そのため、顔に向かうような身振りは、指さしのように直示的な作用を引き起こし、顔に対する視覚的注意を促進したのかもしれない。あるいは、聞き手が身振り情報を理解する過程で「飲む」や「食べる」ということ理解した結果、これらの動作に大きく関わる口元に注意が向けられ、その結果として顔に対する視覚的注意を促進したのかもしれない。また、「食べる」動作の身振りに対する注視時間は成人の方が6歳児よりも長かった。これは、成人が6歳児よりも、身振りから「食べる」に関する具体的な意味情報を抽出しようとしていることを示していたのかもしれない。

**今後の課題** 本研究は、6歳児における発話と身振りの統合能力が成人と同程度ではない可能性を示した。また、その一方で視覚的な注意は6歳児と成人の間で差がみられないことを明らかにした。そのため、発話

と身振りの統合的理解における両者の発達的な差は、認知的な処理過程に起因することが予測される。今後は、話者に対する視覚的注意に加えて、発話と身振りの処理過程に関わる認知能力を検討する必要があるだろう。

また、本研究の参加者は、いずれも発話と身振りの統合能力を有していると想定される者たちであり、6歳児は有意傾向であったものの、おおむね先行研究と同様の結果となった。しかし、両者の統合が明らかに難しいと考えられる6歳児未満の幼児では未検討であった。そのため、今後は、6歳児未満の幼児を対象にして、統合的理解の能力と視覚的注意の関連性について検討する。

最後に、本研究では、身振りを参照する指示語発話を設定しておらず、発話と身振りの統合的理解を促進された場合における聞き手の視線は考慮されていない。そのため、今後は、指示語発話の有無を設定して、発話と身振りの統合的理解と視覚的注意の関連性を検討する必要がある。

## 5. 参考文献

- [1] Kelly, S. D., Özyürek, A., & Maris, E., (2010) "Two sides of the same coin: Speech and gesture mutually interact to enhance comprehension", *Psychological Science*, Vol. 21, No. 2, pp. 260-267.
- [2] Sekine, K., Sowden, H., & Kita, S., (2015) "The development of the ability to semantically integrate information in speech and iconic gesture in comprehension", *Cognitive Science*, Vol. 39, No. 8, pp. 1855-1880.
- [3] Miyake, H., & Sugimura, S., (2018) "The effect of directive words on integrated comprehension of speech and iconic gestures for actions in young children", *Infant and Child Development*, Vol. 27, No. 5, pp. 1-9.
- [4] Watamaki, T., & Ogura, T., (2004) "*Technical manual of the Japanese MacArthur communicative development inventory: Words and grammar*", Kyoto: Kyoto International Social Welfare Exchange Center.
- [5] Cocks, N., Morgan, G., & Kita, S., (2011) "Iconic gesture and speech integration in younger and older adults", *Gesture*, Vol. 11, No. 1, pp. 24-39.
- [6] Cocks, N., Sautin, L., Kita, S., Morgan, G., & Zlotowitz, S., (2009) "Gesture and speech integration: An exploratory study of a man with aphasia", *International Journal of Language and Communication Disorders*, Vol. 44, No. 5, pp. 795-804.