

ジェスチャーと発話の時間的ダイナミクス： 脳磁図による神経基盤の検討

The temporal dynamics between gesture and speech: An MEG Study

関根 和生¹, 大隈玲志¹, 番浩志^{2,3}
Kazuki Sekine, Reiji Ohkuma, Hiroshi Ban

¹早稲田大学, ²情報通信研究機構, ³大阪大学

¹Waseda University, ²Center for Information and Neural Networks, ³The University of Osaka
ksekine@waseda.jp

概要

ジェスチャーが発話産出を促進させることが示唆されてきたが、そのプロセス、特に神経基盤についてはよくわかっていない。本研究では、23名の日本語母語話者を対象に、ジェスチャー産出と抑制の条件下で30秒間のアニメーションを説明させ、MEG（脳磁図）で計測を行った。ジェスチャー使用条件よりも、ジェスチャー抑制条件において、発話開始直前（-0.25～0秒）に脳活動（RMS値）が上昇し、両側前側頭葉の強い活動がみられた。この結果からは、ジェスチャーが発話処理の負荷を軽減させていることが示唆された。

キーワード: ジェスチャー, 発話産出, 自己指向的機能, マルチモーダルコミュニケーション

1. 問題と目的

われわれは、発話とともにしばしばジェスチャーを行っている。発話と共起するジェスチャーは、*co-speech gesture* と呼ばれ、発話と時間的にも意味的にも同期することをその特徴としている。以下では、単にジェスチャーと呼ぶ。ジェスチャーは、自発的な身体活動であり、同じ指示対象を表すときでさえ、文脈や聞き手の属性によって、その形式や動きが変化する。この特徴は、手の意味と形のペアがあらかじめ決められているエンブレム（例：ピース、OKサインなど）とは異なる（McNeill, 1992, 2005）。ジェスチャーはこれまで、主に聞き手への情報伝達手段とみなされてきたが、近年では話し手の発話生成や思考過程に関与する機能にも注目が集まっている（Kita, 2000）。実際に、対面相手が見えない状況（Alibali et al., 2001）や電話越しの会話（野中・関根, 2024）においても話者はジェスチャーを産出しているという事実や、先天的に視覚障害を持つ話者でも自発的なジェスチャーを使用するという報告（Iverson & Goldin-Meadow, 1997）からも、ジェスチャーには視覚的情報の伝達を前提としない認知的役割があると考えられる。

ジェスチャーの自己指向的機能を説明するために、これまでいくつかの産出モデルが提案されてきた。そ

れらはいずれも、Levelt（1989）のモジュール型発話産出モデルを基盤としており、発話産出が *conceptualiser*, *formulator*, *articulator* の段階を経て進行するという情報処理的枠組みに依拠している。たとえば、*Image Activation Hypothesis* (de Ruiter, 1998) は、ジェスチャーが *conceptualiser* の段階において心的イメージを活性化することで、前言語的メッセージの構築を支援すると想定している。*Lexical Retrieval Hypothesis* (Rauscher et al., 1996) は、ジェスチャーが運動的表象として *formulator* の段階に関与し、語検索を促進すると仮定する。また、*Information Packaging Hypothesis* (Kita, 2000) は、ジェスチャーが情報の構造化を担い、発話前の概念形成を助けると位置づけている。

これらのモデルはいずれも、ジェスチャーが発話を促進する認知的機能を理論的に説明しているものの、共通の限界として、ジェスチャーと発話生成の時間的關係についての明確な仮説や実証的根拠を欠いている。その主因として、既存の知見の多くが行動観察に基づくものであり、ジェスチャーの生成がどの時点で言語処理に介入しているのかを明確に捉えるには不十分であったことが挙げられる。発話生成過程が基本的にフィードフォワード型で進行するとされるLeveltの発話産出モデルの枠組み内では、ジェスチャーがリアルタイムで語検索や概念化に寄与するためには、対応する語よりも早期に生成されなければならないが、実際には発話とジェスチャーはしばしば同期して産出されている。この時間的矛盾は、ジェスチャーの実行自体ではなく、ジェスチャー生成の意図や運動準備が発話処理に寄与している可能性を示唆する。

このような理論的・方法的課題を踏まえ、本研究では、時間分解能に優れた神経計測法である脳磁図計測（MEG）を用いて、ジェスチャーが発話に与える影響の時間的経過を明らかにすることを目的とした。

2. 方法

2-1. 実験参加者 日本語を母語とする23名（男性20名、女性3名、 23.2 ± 1.8 歳）が実験に参加した。実験は、情報通信研究機構・人を対象とする研究倫理委員会（承認番号：N2300624090133-FY24-00014）および脳情報通信融合研究センター・安全委員会（2405150120）の承認を得て実施した。全参加者から事前に口頭・書面で同意を取得した。

2-2. 実験計画 本研究は1要因参加者内計画で、独立変数はジェスチャーの産出（産出あり VS. 産出なし）であった。

2-3. 課題 参加者は物語再話課題に取り組んだ。この課題では、約3から9秒のアニメーションを視聴し、その内容を口頭で説明することが求められた。

2-4. 手続き 課題はジェスチャー産出条件 (Encouraged gestures condition) とジェスチャー抑制条件 (Restrained gestures condition) の二条件下で実施された。参加者は脳磁計ブースの中に入り、腹部前に固定された机に両手を置きながら、座位の姿勢を維持した。物語再話課題に取り組むため、眼の前に設置されたスクリーンに投影されるアニメーションを視聴した。動画終了後に数秒考える時間が与えられた。その後、スクリーン上にプロンプトが出現し、10秒程度でアニメーションの内容を説明した。その際、ジェスチャー産出条件では明示的にジェスチャーの使用が奨励された。肘は机に付けたままで産出することが求められた。一方、ジェスチャー抑制条件では参加者は手をテーブルに置いた状態で発話するよう指示された。1試行の時間は約30秒程度である。1回の撮像セッションで30回試行を行い、各条件（ジェスチャー産出・抑制）につき、3から4セッションの計測を行った。参加者一人あたりの総試行数は90-120試行となった。2つの条件の実施順はカウンターバランスを取った。

2-5. MEG計測 脳活動は、Elekta Neuromag 360 (MEGIN Oy, Helsinki, Finland) を用いて1000Hzで記録された。

2-6. MEG解析 発話やジェスチャーの開始・終了時間は、3名のコーダーが実験時に録画されたビデオからアノテーションソフトELANを用いて同定した。MEGデータはOversampled temporal projection (OTP), XSCAN, Spatiotemporal signal space separation (tSSS)を用いて前処理された。さらに、Notch Filter, Down Sampling(200Hz), Band Pass

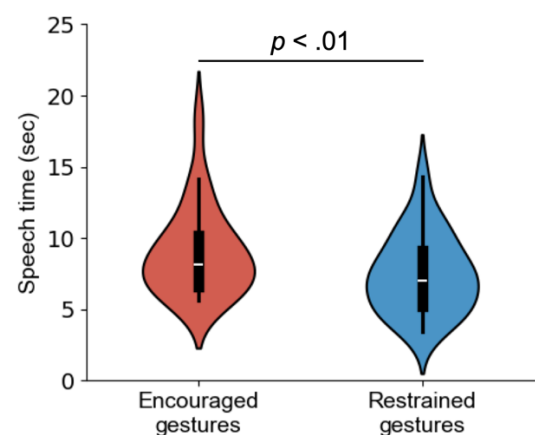
Filter (0.1-45Hz), ICA と不要成分除去(眼電, 心電, 筋電), Global Mean によるベースライン補正によって、生体由来ノイズが除去された。

前処理後のMEGデータは、発話開始を基準(0秒)にアラインされた。続いて、脳活動全体の傾向を把握するため、全チャンネルのRoot Mean Square (RMS)を求めた。さらに、発話産出に関与する左前側頭葉(Left ATL)と右前側頭葉(Right ATL), 運動に関与する運動野(Motor Area)の事象関連活動を求めた。RMSより条件間で差があることが示唆された。0.25秒から0秒の平均振幅を用いて、対応あり t 検定を行った($\alpha=0.05$)。

3. 結果

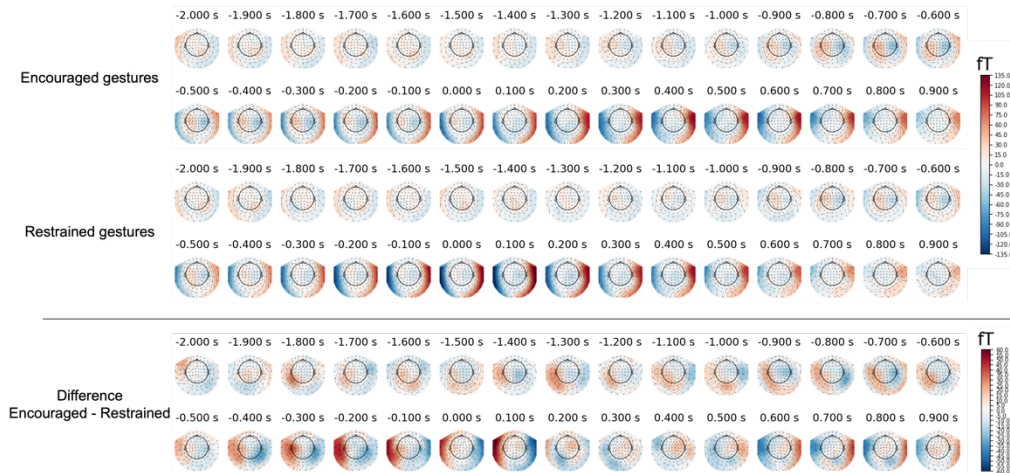
3-1. 発話時間の比較 ジェスチャーが発話に与える影響を検討するため、1試行あたりの平均発話時を計測した。図1は、ジェスチャー使用を促す条件 (encouraged gestures condition) およびジェスチャーを抑制する条件 (restrained gestures condition) における発話時間を示している。対応のある t 検定を行ったところ、両条件間で有意な差が認められ、ジェスチャーを使用した場合の方が、使用を抑制した場合よりも発話時間が有意に長かった ($t(22)=3.43, p=.002$)。

図1. 両条件における発話時間 (秒)



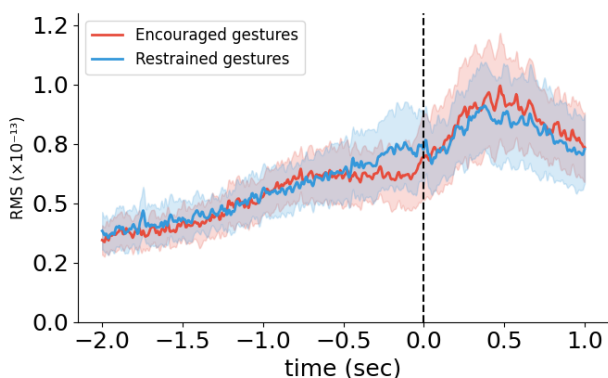
3-2. 脳活動の比較 発話開始 (0秒) を基準として、-2秒から+1秒までの電位分布を0.1秒ごとに算出した。図2は、ジェスチャー使用条件と抑制条件における脳活動の時間推移をトポグラフィマップで示している。差分マップから、両条件間の神経活動における乖離が-0.3秒から+0.1秒の間で生じていることが示唆された。

図 2. 発話開始前後（-2～+1 秒）のトポマップの時間推移。0 秒は発話開始を示す。両条件間の違いは-0.25 秒～0 秒の間に観察された。上段：ジェスチャー使用条件，中段：ジェスチャー抑制条件，下段：両条件の差分。なお，MEG 装置は立体的なヘルメット型構造をしており，2 次元平面への投影により一部のチャンネルが頭部の外に見える場合がある。



両条件間の差が生じた時間窓をさらに特定するために，すべてのチャンネルにおける RMS (Root Mean Square) 値の合計を算出した (図 3)．その結果，-0.25 秒から 0 秒の範囲において両条件の差が明確に確認された．したがって，以降の分析ではこの時間窓に焦点を当てることとした．

図 3. 発話開始にアラインされた RMS 波形．RMS 値に基づき，ジェスチャー使用条件と抑制条件の差が-0.25 秒～0 秒の範囲で観察された．統計解析は実施していない．色付き部分は 95%信頼区間を示す．



RMS の差異は言語処理と関係の深い側頭葉の活動の違いを反映している可能性があることから，関心領域 (ROI) として両側の前側頭葉に着目し，さらにジェスチャーや発話に関連する運動活動による影響を排除するため，運動領域も対象に含めて分析を行った．その結果，-0.25～0 秒の間，両側前側頭葉における神経活動は，ジェスチャー使用条件において抑制条件より

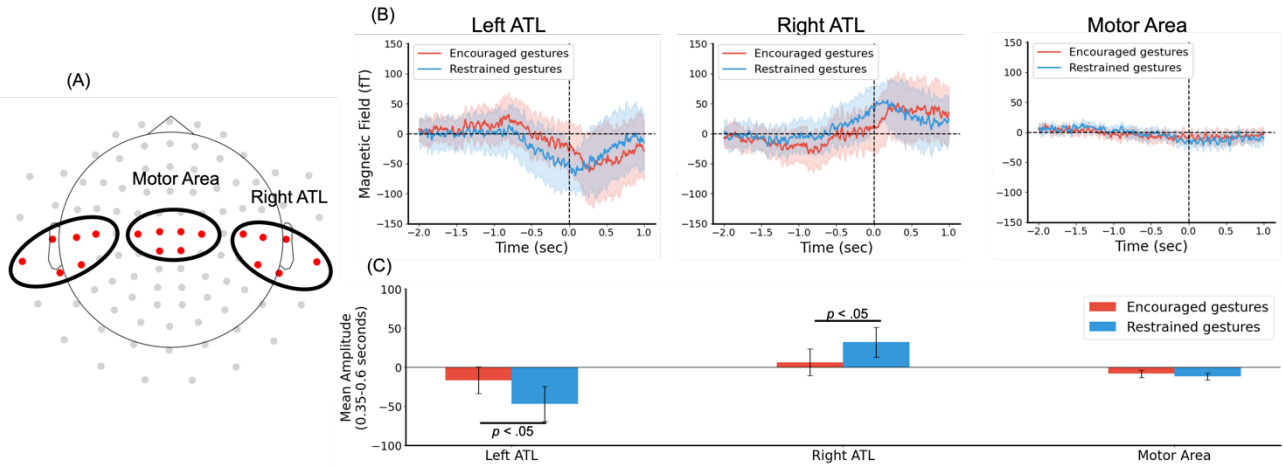
りも低かった (図 4)．図 4A は右前側頭葉，左前側頭葉，および運動領域に対応するチャンネルを示し，図 4B は各 ROI における平均波形，図 4C は同時間窓内の平均振幅を示している．発話開始前後に両側前側頭葉では明確な事象関連磁場 (ERFs) が観察されたが，運動領域では明確な ERFs は見られなかった．対応のある t 検定の結果，ジェスチャー抑制条件における前側頭葉の活動は，使用条件よりも有意に高かった (左前側頭葉： $t = 2.08, p = .049$ ；右前側頭葉： $t = 2.35, p = .028$)．一方，運動領域では有意な差は認められなかった ($t = 1.08, p = .291$)．

4. 考察

本研究は，時間分解能に優れた脳磁図計測 (MEG) を用いて，ジェスチャーが発話産出に与える影響の時間的側面を明らかにすることを目的とした．その結果，発話開始 250 ミリ秒前に，両側前側頭葉 (anterior temporal lobes; ATL) の神経活動が，ジェスチャー使用条件において有意に抑制されることが示された．一方，運動関連領域には有意差が認められなかった．このことから，ジェスチャーの効果は一般的な運動準備によるものではなく，言語処理に特異的であると考えられる．

これらの知見は，語検索の過程においてジェスチャーが意味的表象を活性化し，それによって語彙アクセスを支援するという Lexical Retrieval Hypothesis (Rauscher et al., 1996) を神経生理学的に支持するものである．前側頭葉は意味記憶と概念処理の中核を担う領域であり，

図 4. ROI ごとの事象関連磁場 (ERFs)。(A) 各関心領域 (ROI) に対応するチャンネルの定義。(B) 各 ROI における平均波形。左前側頭葉 (左), 右前側頭葉 (中央), 運動領域 (右)。(C) -0.25~0 秒間の平均振幅。0 秒は発話開始点を示す。パネル B の陰影部分は 95%信頼区間, パネル C のエラーバーは標準偏差を示す。



特に語義の検索や意味的曖昧性の解消に関与するとされている (Lambon Ralph et al., 2017). 本研究における前側頭葉の活動の低下は, ジェスチャーによって意味処理の負荷が軽減された結果である可能性が高い。

加えて, 本研究の結果は, ジェスチャーが発話前の概念的情報を整理・構造化する役割を果たすとする Information Packaging Hypothesis (Kita, 2000) や, 心的イメージの活性化を通じて発話準備を促進する Image Activation Hypothesis (de Ruiter, 1998) とも整合する。従来の行動実験では, ジェスチャーの効果は主に発話内容や流暢性に基づいて間接的に推測されていたが, 本研究は MEG を用いて, 発話前段階におけるジェスチャーの神経学的効果を直接的に示した点で新規性がある。

さらに, 本研究はジェスチャー研究における方法論的ギャップも埋めるものである。先行研究の多くは fMRI や fNIRS を用いて空間的局在に焦点を当てていたが, 本研究は MEG の時間精度を活かすことで, ジェスチャーが発話準備過程においてどの時点で言語処理に影響を与えるかを明示的に捉えた。これにより, ジェスチャーが単なる表現手段ではなく, 発話生成過程に統合された認知的・言語的モダリティであることが示唆された。

文献

Alibali, M. W., Heath, D. C., & Myers, H. J. (2001). Effects of visibility between speaker and listener on gesture production: Some gestures are meant to be seen. *Journal of Memory and Language*, 44(2), 169–188.

de Ruiter, J. P. (1998). Gesture and speech production. *Unpublished doctoral dissertation*, Nijmegen University, The Netherlands.

Iverson, J. M., & Goldin-Meadow, S. (1997). What's communication got to do with it? Gesture in blind and sighted children. *Developmental Psychology*, 33(3), 453–467.

Kita, S. (2000). How representational gestures help speaking. In D. McNeill (Ed.), *Language and gesture* (pp. 162–185). Cambridge University Press.

Lambon Ralph, M. A., Jeffries, E., Patterson, K., & Rogers, T. T. (2017). The neural and computational bases of semantic cognition. *Nature Reviews Neuroscience*, 18(1), 42–55.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. The MIT Press.

McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. University of Chicago Press.

McNeill, D. (2005). *Gesture and thought*. University of Chicago Press.

野中郁子・関根和生. (2024). 電話応対時のお辞儀が音響特性に与える影響 日本認知科学会第 41 回大会

Rauscher, F. H., Krauss, R. M., & Chen, Y. (1996). Gesture, speech, and lexical access: The role of lexical movements in speech production. *Psychological Science*, 7(4), 226–231.