不一致な co-speech gesture は発話に明示されない時点情報を補うのか？

Do incongruent co-speech gestures supplement time-related information that is not explicitly expressed in speech?

小林 春美[†]，浅野 知仁[†]，安田 哲也[‡]
Harumi Kobayashi, Tomohito Asano, Tetsuya Yasuda
[†]東京電機大学，[‡]山口大学
Tokyo Denki University, Yamaguchi University
h-koba@mail.dendai.ac.jp, tetsu-yasuda@yamaguchi-u.ac.jp

概要

This study examined whether incongruent co-speech gestures help listeners infer time-related information that is not explicitly expressed in speech. Twenty Japanese adults viewed short videos in which speech and gesture were either congruent or incongruent, while speech referred to either the present or the future. After each video, participants judged what the speaker wanted “now” and “next” from multiple alternatives. We predicted that, in incongruent trials, listeners would rely on speech for the time point explicitly mentioned in speech, but would use gesture to infer information about the unmentioned time point. The results supported this prediction: when speech and gesture were incongruent, participants tended to interpret the time point not explicitly expressed in speech on the basis of gesture. In contrast, when speech and gesture were congruent, speech-based interpretations remained dominant. These findings suggest that incongruent co-speech gestures can supplement temporally unexpressed information in utterance comprehension.

キーワード: **発話随伴ジェスチャー (co-speech gesture) 発話とジェスチャーの不一致, 時点情報**

1. はじめに

人は会話において，発話に伴って自発的にジェスチャーを産出する。こうした co-speech gesture は，単なる付随的な身体動作ではなく，発話と密接に結びつきながら情報伝達に関与する。実際，ジェスチャーはしばしば発話だけでは十分に表されない内容を担い，聞き手の理解を支える手がかりとなることが指摘されている（喜多, 2000; Goldin-Meadow & Wagner, 2005）。

一方で，発話に伴うジェスチャーは，常に発話と同じ内容を示すとは限らない。Goldin-Meadow and Wagner (2005) は，発話内容とジェスチャーの不一致について，ジェスチャーが発話には含まれない情報を担う現象として論じている。たとえば，ピアジェの量の保存課題において，発話では容器の直径に言及しないが，手指では容器の直径を示すような子どもは，高さだけでなく直径も考慮すべきだと理解しつつあり，認知的に移行状態にあると推測できる。

一方，McNeill (1987, 2015)の成長点仮説では，ジェスチャーと発話は共通の表象的基盤から生起し，前者はイメージ的側面を，後者は言語的側面をより強く担うと想定している。この考えに立つと，発話とジェスチャーの不一致は，異なる情報が両モダリティに分配されて現れた状態として捉えることが可能である。日常生活においては，この分配が時間をまたいで起きる場合も観察される。著者のうちの一人が遭遇した事例では，レストランで話者が「今は水だけください」と言いながら，指さす先がメニューのデザートの写真だった場合があった。話者は頭の中で「とりあえず水を頼み，デザートは後で頼もう」と考えているため，発話とジェスチャーがずれた状態になったと考えられる。それでも観察者はこの様子から，「今は水だけ頼むが後でデザートを頼むのだろう」と推測可能である。

本研究では，ジェスチャーと発話内容の不一致を，単なる誤りや認知的移行状態の指標として見るのではなく，発話とジェスチャーに時間的に異なる情報が分配されたとき，聞き手が発話に明示されない時点情報をジェスチャーから補うかという観点から検討する。

とくに本研究で着目するのは，「今」や「次」といった時点情報である。たとえば，「今欲しいものはハサミです」と言いながら包丁で切るようなジェスチャーを行う場合，聞き手は発話に示された現在の情報（今欲しいものはハサミ）に加えて，発話には明示されていない後続時点の情報（次に欲しいものは包丁）をジェスチャーから補って解釈する可能性がある。すなわち，本研究では，不一致な co-speech gesture が，発話に明示されない時点情報を担いうるかを，聞き手の解釈から検討する。

以上を踏まえ，本研究では，発話が現在または未来の

どちらの時点を示しているかを操作したうえで，発話とジェスチャーの一致・不一致が聞き手の解釈に及ぼす影響を検討した。具体的には，発話に明示された時点の情報だけでなく，発話には明示されていない別時点の情報について，聞き手がどの程度ジェスチャーを手がかりとして用いるかを調べた。不一致条件では，聞き手は，発話に示された時点では発話を基準に解釈する一方，発話に明示されていない時点については，ジェスチャーを手がかりとして補完的に解釈すると予測した。これに対し，一致条件では，両時点において発話とジェスチャーが同じ内容の情報を与えるため，全体として発話基準の解釈が優勢になると予測した。

2. 方法

実験参加者は理工系大学に在籍する 19 歳から 26 歳の学部生，大学院生 20 名であった。

実験では，発話とジェスチャーの一致・不一致と，発話が示す時点（現在・未来）を操作した映像刺激を用いた。具体的には，不一致・現在，不一致・未来，一致・現在，一致・未来の 4 条件を設定し，各条件 6 項目，計 24 項目を作成した。参加者は各映像を視聴した後，話し手が「今」および「次」に欲しているものをどのように解釈するかを回答した。

刺激は，話し手が欲しいものを発話とジェスチャーで表現する短い映像であった。不一致条件では，たとえば「今欲しいものはハサミです」と発話しながら包丁で切るジェスチャーを行うように，発話とジェスチャーが異なる対象を示した。一致条件では，「今欲しいものはハサミです」と発話しながらハサミで切るジェスチャーを行うように，発話とジェスチャーが同じ対象を示した。

各映像の後，参加者は「この人物が今欲しいものは何か」「この人物が次に欲しいものは何か」という 2 つの質問に答えた。回答は複数の選択肢から行い，不一致条件では発話とジェスチャーが不一致であり，選択肢は発話項目，ジェスチャー項目，カテゴリー類似項目，ジェスチャー類似項目を含めた。一致条件では発話とジェスチャーが同一であり，同様の項目から成る選択肢を設けた。

実験は対面で 1 名ずつ実施した。参加者はパソコン上で 2〜3 秒程度の映像刺激を視聴し，各試行後に回答用紙へ記入した。実験前には，実験者は参加者に対して，映像の音声と動作をよく見聞きすること，映像は必要に応じて再生可能であること，加えて「今」は映像の最中，「次」はそれより後の時点を指すことを説明した。刺激系列には 2 種類の提示順を用意し，参加者には両系列が同数ずつ割り当てられた。

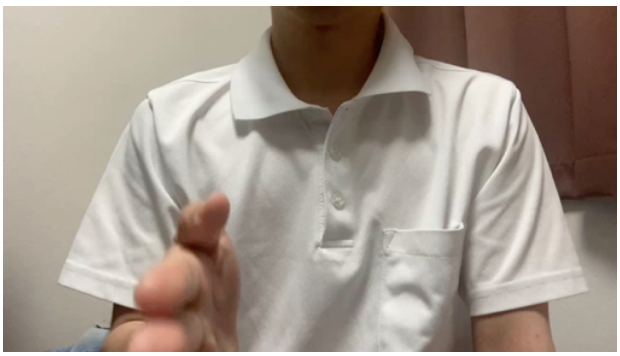

図 1　実験で提示した映像刺激．上半身が提示されたが表情等は見えなかった．

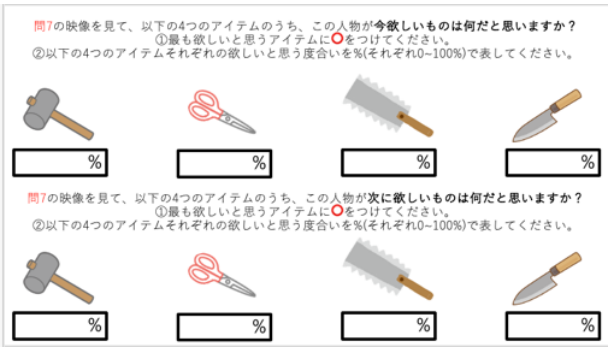

図 2　不一致条件における選択肢の例

統計モデリングには R ソフトウェア(R Core Team, 2024)を用いた。使用したパッケージは線形混合モデルを構築するための lme4(Bates et al., 2015)，統計値を産出するための lmerTest パッケージ (Kuznetsova et al., 2017)，データハンドリングを行うための tidyverse パッケージ (Wickham et al., 2019)，推定値を出力するための ggeffects パッケージ(Lüdecke, 2018)を用いた。

分析には，参加者が発話とジェスチャーのどちらを選択したのかを調べるために，擬似的ロジット変換 (empirical logit: Barr, 2008)を用いた従属変数を使用し，参加者の選択傾向を調べられるようにした。この従属変数を利用し，線形混合モデル（LMM）を用いた分析を行った。LMM は，lme4 パッケージの lmer 関数を用

いて構築した。まず，実験条件とその交互作用を固定効果として含む最大モデルを構築した。その後，得られたデータにフィットするようなモデルの候補を調べ，モデルを選定した。分析モデルは，最大モデルと同様のモデルが選定された。

3. 結果

LMM の結果，Intercept (β = -0.25, t = -3.91, p < .001)，発話情報(今，次)と質問文(今欲しいものは何か，次に欲しいものは何か)の交互作用(β=-2.50, t=-17.63, p<.001)が認められた。しかしながら，発話情報（β=-0.01, t=-0.16, p=0.88），質問文(β= 0.05, t= 0.71,p=0.48)に関しての違いはなかった。

交互作用を調べるために，emmeans パッケージ(Lenth & Piaskowski, 2026)を用い，単純主効果を調べた。質問「今欲しいものは何か」では，「今」について言っている場合 (EMM = -0.89, 95%CI[-1.01, -0.77])の方が「次」について言っている場合 (EMM = 0.35, 95%CI[0.15, 0.55])よりも発話基準で解釈していた(OR = -1.24, t = -11.58 , p < .05)。

質問「次に欲しいものは何か」では，「今」について言っている場合 (EMM = 0.41, 95%CI[0.22, 0.60])の方が「次」について言っている場合 (EMM = -0.85, 95%CI[-1.22, -0.48])よりもジェスチャー基準に解釈していた(OR = 1.26, t = 11.81 , p < .05)。

「今」について言っている場合では，質問「今欲しいものは何か」(EMM = -0.89, 95%CI[-1.01, -0.77])の方が質問「次に欲しいものは何か」(EMM = 0.41, 95%CI[0.22, 0.60])よりも発話基準に解釈していた(OR = -1.30, t= -12.97 , p < .05)。「次」について言っている場合では，質問「今欲しいものは何か」(EMM = 0.35, 95%CI[0.15, 0.55])の方が質問「次に欲しいものは何か」(EMM = -0.85, 95%CI[-1.22, -0.48)よりもジェスチャー基準に解釈していた(OR = 1.20, t = 11.96 , p < .05)。

つまり，「今欲しいものはハサミです」と言いながら包丁のジェスチャーをしている人に関して，「この人が今欲しいものは何ですか？」と質問されると，参加者は，発話を基準として「ハサミ」と答えやすい。一方，「この人が次に欲しいものは何ですか？」と質問されると，参加者は，ジェスチャーを基準として「包丁」と答えやすい。

これに対し，一致条件でも交互作用は認められたものの（β = -0.57, t = -9.28, p < .001），全体としては発話項目が優勢であった。したがって，発話とジェスチャーが一致している場合には，聞き手は主として発話に依拠して解釈しており，不一致条件で見られたようなジェスチャーによる補完は限定的であった。

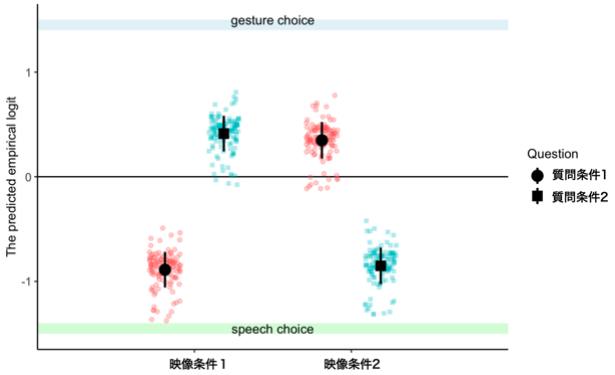

図 3　不一致条件における結果．各値は LMM から出力した．質問条件 1 : 今欲しいものは何か，質問条件 2 : 次に欲しいものは何か，映像条件 1 : 今について言っている映像，映像条件 2 : 次について言っている映像，をそれぞれ示す。

4. 考察

本研究の結果は，co-speech gesture が単に発話内容を重複して伝えるだけではなく，発話に明示されていない別時点の情報を補完する役割を担いうることを示している。不一致条件では，「今」と「次」のいずれを発話が示しているかにかかわらず，聞き手は発話に含まれない時点の情報についてジェスチャーを参照していた。これは，聞き手が発話とジェスチャーを相互に無関係な信号として処理するのではなく，両者を統合しながら話し手の意図を推定していることを示唆する。

この結果は，発話とジェスチャーの統合的生成を想定する McNeill の議論や，理解段階における相互作用を示した Kelly らの知見(Kelly et al., 2010)と整合的である。とくに本研究は，ジェスチャーの役割を「対象の補助的説明」としてではなく，「時間的にずれた情報の橋渡し」として捉え直す点に意義がある。話し手が現在のことを発話で述べつつ，別の時点に関わる内容をジェスチャーで示した場合でも，聞き手はそのずれを無視するのではなく，むしろ情報統合の手がかりとして利用していた可能性がある。

もっとも，ジェスチャーは常に発話と同等の強さで解釈を導いたわけではなかった。この点は，刺激が短い

映像であり，実際の会話に比べて文脈や相互行為性が限られていたことと関係しているかもしれない。また，ジェスチャーの種類によっては，意図した対象が十分に伝わらなかった可能性もある。今後は，より自然な会話場面を用いること，ジェスチャーの視認性や慣習性の違いを検討することによって，どのような条件でジェスチャーが時間情報の補完に寄与するのかをさらに明らかにする必要がある。

本研究では，発話とジェスチャーの一致・不一致および発話のタイミング情報を操作し，聞き手の解釈を検討した。その結果，発話と不一致な co-speech gesture は，発話に明示されていない時点の情報を補う手がかりとして機能することが示された。co-speech gesture は，発話の補助にとどまらず，時間的に分散した情報を統合する役割をもつ可能性がある。

謝辞

実験に参加くださったすべての参加者に感謝いたします。この研究は科研費 23K02901(HK) と 25K07663(TY)の補助を受けて行いました。

文献

Barr, D. J. (2008). Analyzing “visual world” eyetracking data using multilevel logistic regression. *Journal of Memory and Language, 59*(4), 457–474.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. Journal of Statistical Software, 67, 1-48.

Goldin-Meadow, S., & Wagner, S. M. (2005). How our hands help us learn. Trends in Cognitive Sciences, 9(5), 234–241. https://doi.org/10.1016/j.tics.2005.03.006

Kelly, S. D., Özyürek, A., & Maris, E. (2010). Two sides of the same coin: Speech and gesture mutually interact to enhance comprehension. *Psychological science*, *21*(2), 260-267.

Kuznetsova, A., Brockhoff, P.B., Christensen, R.H.B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, **82**(13), 1–26.

Lenth, R., & Piaskowski, J. (2026). emmeans: Estimated Marginal Means, aka Least-Squares Means. R package version 2.0.3, https://rvlenth.github.io/emmeans/.

Lüdecke, D. (2018). ggeffects: Tidy Data Frames of Marginal Effects from Regression Models. *Journal of Open Source Software*, 3(26), 772.

喜多壮太朗 (2000). 人はなぜジェスチャーをするのか. 認知科学, 7(1), 9-21.

McNeill, D. (1987). Psycholinguistics. New York: Harper & Row.

McNeill, D. (2015). Why we gesture: The surprising role of hand movements in communication, Cambridge University Press.