

# Interplay of speech and spontaneous manual gestures

Anshul Bawa, Sumeet Agarwal, Rajakrishnan Rajkumar  
IIT Delhi

## Introduction

- Interpersonal face-to-face interaction often involves visible hand movements called gestures.
- There exists an intrinsic semantic and temporal relationship between spontaneous speech and co-occurring gestures, and there is evidence of common cognitive processes and areas involved in the production of both.
- Using the Speech and Gesture Alignment (SaGA) corpus, we investigate the semantics and pragmatics of different kinds of gestures, and provide empirical answers to the debate between whether the gesture parallels or compensates for speech, in terms of descriptivity and articulation.

## Functional Classification of Gestures

1. emblems/symbols, like in ``Hello" and ``Bye",
2. iconic gestures, to depict actions, objects or space,
3. deictic gestures, to point to objects or in a direction,
4. beat gestures, to coordinate or emphasise speech, and
5. gestures serving a discourse function.

## Contrasting Hypotheses

Because gestures are be both redundant and non-redundant, there are different perspectives on modelling the interaction between speech and gesture.

- The Tradeoff Hypothesis suggests that the two modalities complement each other, and that they compete for production, with that modality being favoured which is easier to articulate in that context.
- This account is conflicted by another hypothesis called the Hand-in-Hand Hypothesis. Speakers' gestures were found to parallel underspecifications in speech, and it is deduced that both are planned simultaneously.

## Method

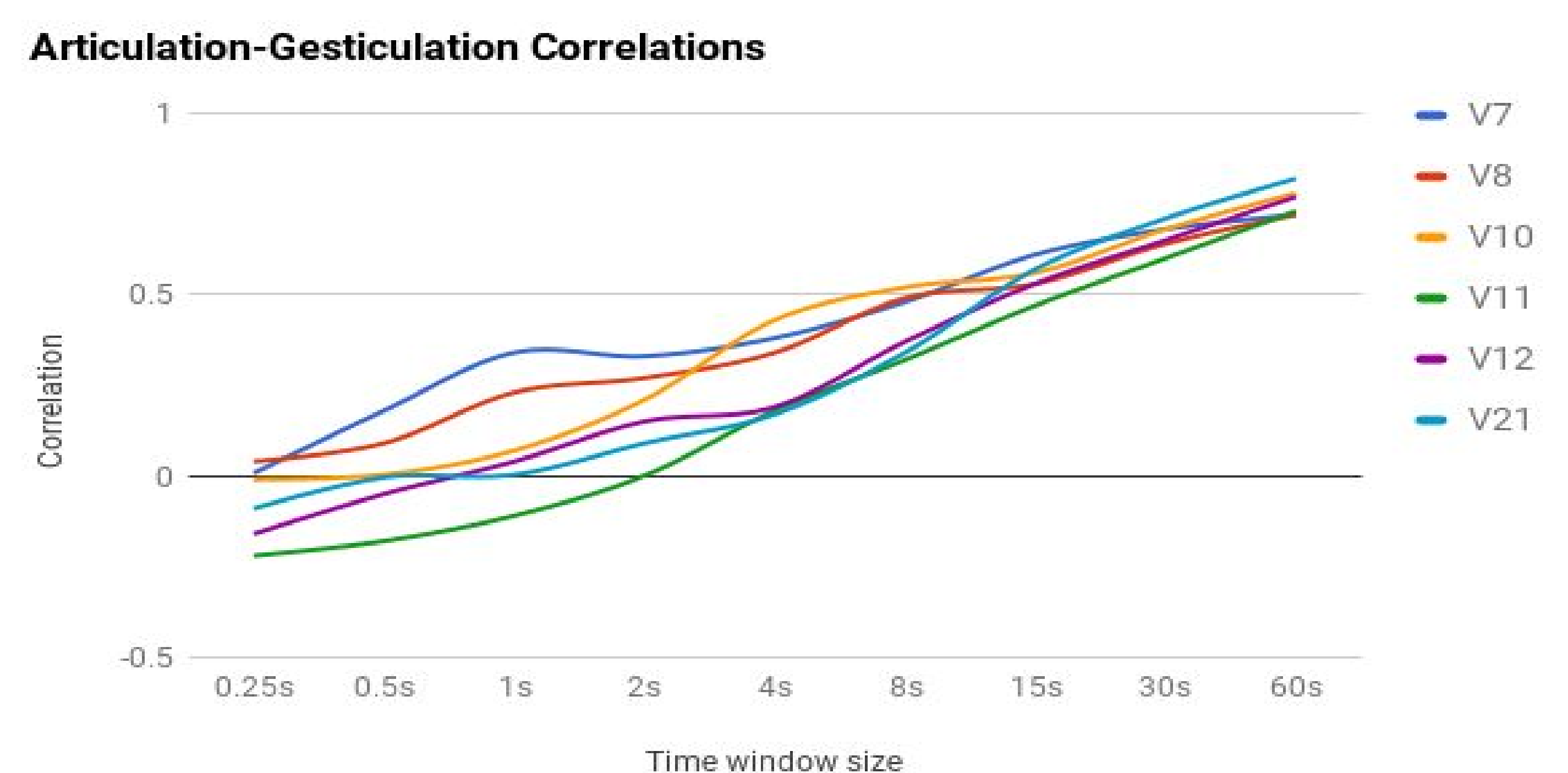
- The rates of production are operationalized as number of words spoken, number of discrete gesture events, and time spent speaking/gesturing.
- Speech transcriptions, and annotations about the referents (landmarks, navigational actions, etc.) are used to extract semantic features.
- Semantic features, distributions of words and past gestures are used to predict a gesturing event and its type through a one-hidden layer neural network. Multivariate regression analysis is used to infer rate correlations.

## References

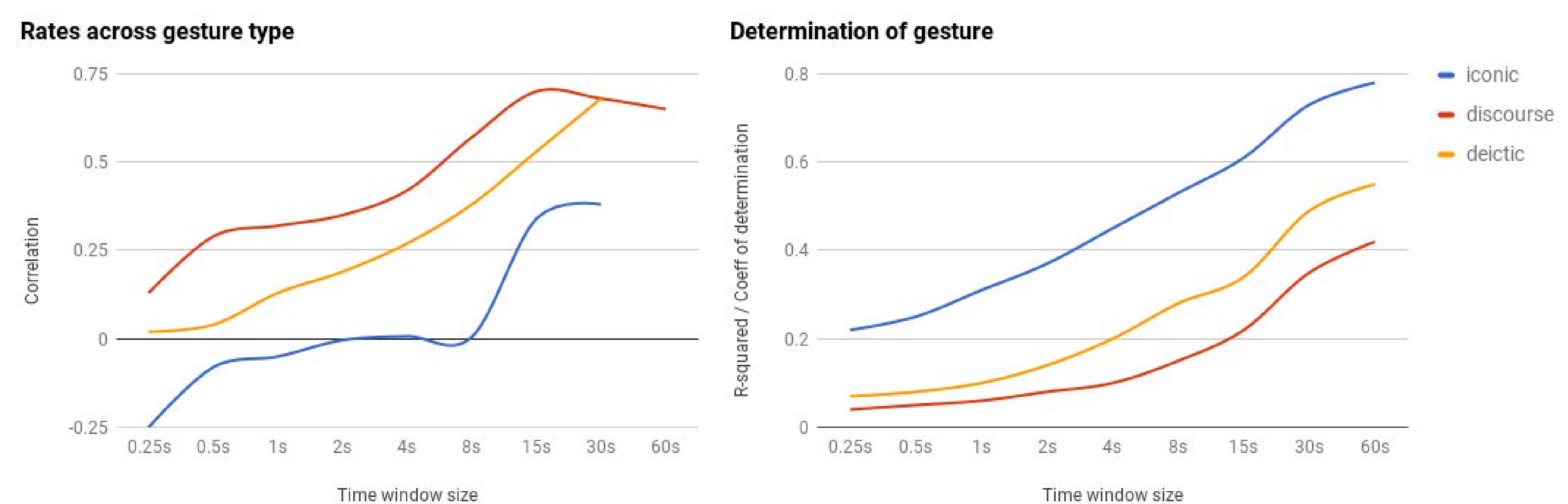
- Jan P De Ruiter, Adrian Bangerter, and Paula Dings. 2012. The interplay between gesture and speech in the production of referring expressions: Investigating the tradeoff hypothesis. Topics in Cognitive Science 4(2):232–248.
- Asli Özyürek, Roel M Willems, Sotaro Kita, and Peter Hagoort. 2007. On-line integration of semantic information from speech and gesture: Insights from event-related brain potentials. Journal of cognitive neuroscience 19(4):605–616.
- Wing Chee So, Sotaro Kita, and Susan Goldin-Meadow. 2009. Using the hands to identify who does what to whom: Gesture and speech go hand-in-hand. Cognitive Science 33(1):115–125.

## Results

For different gesture types (iconic, deictic and discourse) the correlation between articulation of speech and gesture varies consistently across subjects (N = 6).



The more semantically predictable a gesturing event is from its co-occurring speech, the more likely it is to interfere with production rate. Iconic gestures are most easily determined from speech referents (such as action-oriented utterances ( $R^2 = 0.14$ ) and distinctive physical attributes of landmarks ( $R^2 = 0.18$ )) and show weak negative correlation with speech articulation ( $r = -0.15$ ).



Discourse gestures are the least predictable from speech, in the dataset, yet show a moderate positive articulatory correlation ( $r = +0.32$ ). Deictic gesturing (pointing or indexing) occur mostly independent of articulation rates.

## Discussion

It is found that iconic gestures compensate more (tradeoff) whereas discourse gestures parallel speech more (hand-in-hand) from the standpoint of articulation fluency. Since iconic gestures have a referential alignment with speech, they can be thought to share communicative load from the other modality - this is indicative of a co-competitive production mechanism. Though conversational speech does contain many discourse markers, and it is possible that they faithfully synchronize with discourse gestures, but the data annotation does not bring out this descriptive overlap. This difference is consistent with brain imaging studies that show a much higher overlap of iconic gestures with speech, than other gesture types.