

Probabilistic relation between co-speech gestures, pitch accents and information status

Suyeon Im & Stefan Baumann*

Abstract. This study investigates the occurrence of co-speech gestures as a function of prosodic prominence (pitch accents) and discourse meaning (information status) in a clear and engaging speech style. Among several types of co-speech gestures, we examine non-referential gestures, which are claimed to be prosodic in nature (Shattuck-Hufnagel & Ren 2018). In particular, we want to find out to what extent these gestures co-occur with specific accent types and whether they are used to encode referential, lexical, or contrastive information. Our results show that the occurrence of gestures was highest for L+H*, followed by H*, !H*, and unaccented words. Gestures were accompanied by L* only in continuations. Also, co-speech gestures were more likely to occur with new or accessible, and especially contrastive, information than with given information. The patterns differed between the referential and lexical level of information status, though. In general, this study suggests that co-speech gestures contribute to the probabilistic encoding of a word's information status in conjunction with pitch accents.

Keywords. co-speech gestures; prosody; pitch accents; information status; RefLex Scheme; clear speech

1. Introduction. Gestures often co-occur with speech in communication. Co-speech gestures include any visible movement of the body including the hands, the arms, the head, and the eyebrows (Kendon 2004). Hand gestures can be described in terms of their functions in communication, such as emblematic, iconic, deictic, and beat gestures (McNeill 1992). These gestures have been found to be temporally aligned with prosodic prominence in prior empirical studies (Jannedy & Mendoza-Denton 2005, Loehr 2012, Shattuck-Hufnagel et al. 2007, McClave 1994, Yassinik et al. 2004, among many others). Furthermore, both co-speech gestures and prominence have been shown to mark the semantic or pragmatic meaning of a word in discourse context (de Ruiter et al. 2010, Holler & Stevens 2009, Kendon 2004, Loehr 2012, McNeill 1992, among many others). Considering their close relations with co-speech gestures, prosodic prominence and discourse meaning will be introduced in more detail in the following section.

Prominent words are higher in pitch, longer in duration, or louder in intensity relative to surrounding words in an utterance. Among the acoustic correlates, pitch accents are described in terms of pitch contour such as high tones (H*), low tones (L*), downstepped high tones (!H*), rising tones (L+H*), etc. (Veilleux et al. 2006). The perception of prominence varies with the type of pitch accent (Bishop et al. in press, Hualde et al. 2016): L+H* is more likely to be perceived as prominent than H*, which in turn tends to be perceived as more prominent than !H*. L* can be perceived as prominent in the case of yes-no questions or the continuation pattern (in combination with a H boundary tone). These pitch accents can be arranged in increasing order of perceived prominence: L* < !H* < H* < L+H*. That is, L* is the least prominent, and L+H* is the most prominent accent type. These ranks of pitch accents will be referred to as the *Prominence hierarchy* in this study.

* Authors: Suyeon Im, Hanyang University (suyeonim03@gmail.com) & Stefan Baumann, University of Cologne (stefan.baumann@uni-koeln.de).

Information status (IS) refers to the cognitively active status of a word in discourse (Chafe 1976, 1994, Lambrecht 1994). A binary distinction of information status (*given* versus *new* information) can be insufficient to capture multiple layers of discourse meaning. Among many possible labels, we introduce four IS labels in this study, namely *given*, *bridging*, *unused*, and *new* (Prince 1981). *Given* refers to the items that are known to hearers or present in context (e.g., repeated items). *Bridging* indicates the items that are referable from the previous context (e.g., whole-part relation). *Unused* describes the items that are assumed to be known to hearers but that are not mentioned in context (e.g., the name of a person or city). *New* denotes the items that are assumed to be unknown to hearers in context. These four IS labels can be arranged in increasing order of newness: *given* < *bridging* < *unused* < *new*. This order of IS labels will be referred to as the *Information status hierarchy* in this study. In addition to these four IS labels, we present three levels of IS: referential, lexical, and contrastive information (Schwarzschild 1999, Halliday & Hasan 1976, Halliday & Matthiessen 2004, Rooth 1992). The referential level describes the status of referring expressions in discourse context. The lexical level depicts the cognitive activation status of lexical expressions or concepts. The “alternative” level, adopted from Rooth’s (1992) notion of focus as relating the meaning of an expression to a set of alternatives (hence the label *alt*, see below), captures the relationship between referring expressions in contrast.

There are few studies on the relation among co-speech gestures, accent types, and information status. For the relation between accent types and information status, it was claimed that there is a (near) one-to-one mapping (Pierrehumbert & Hirschberg 1990). In Table 1, accent types and information status, aligned with the Prominence hierarchy and the Information status hierarchy, respectively, are mapped to each other: unaccented words or L* with given information, !H* with bridging information, H* with unused information, and L+H* with new information. More recent empirical studies, however, suggest that pitch accents probabilistically encode information status (Baumann & Riester 2013, Cangemi & Grice 2016, Im et al. 2018).

| | | | |
|----------------|----------|--------|------|
| Unaccented, L* | !H* | H* | L+H* |
| ↑ | ↓ | ↑ | ↓ |
| Given | Bridging | Unused | New |

Table 1. Mapping between pitch accents and information status

For the relation between co-speech gestures and accent types, most prior studies examined their temporal alignment (Jannedy & Mendoza-Denton 2005, Loehr 2012, Shattuck-Hufnagel et al. 2007, McClave 1994, Yassinik et al. 2004, among many others). It is still unknown how the occurrence of gestures varies with the strength of perceived prominence (i.e., accent types). For the relation between co-speech gestures and information status, a few empirical studies exist (Loehr 2004, 2012, Bergmann & Kopp 2006). Bergmann and Kopp (2006) investigated gesture occurrence in relation to the semantic meaning of a word in a corpus of map description tasks. Results showed that the occurrence of complementary gestures was higher for new entities than for given ones. Redundant gestures, however, showed the opposite pattern in that the occurrence was more frequent for given entities than for new entities. Taken together, the findings suggest that speech was aided by complementary, but not redundant, gestures for the words requiring the speaker’s high cognitive load (i.e., new information). Loehr (2012), in a corpus of casual conversations between friends, made qualitative observations of some gestures as a function of information status, as well as accent types. For new information marked by H*, speakers in the corpus

adopted abstract deictic gestures. Focused expressions were delivered by H* and deictic gestures combined with beats. Emphasis was conveyed by H* or L+H*, as well as by emphatic beats. Although these prior studies yield fruitful insights for our understanding, they are limited in the temporal coordination between gestures and prominence, and they only covered the relation between gestures and a few types of information status. It is still unclear how the occurrence of gestures varies with different types of pitch accents and multiple layers of information status.

The present study investigates the occurrence of co-speech gestures in relation to accent types and information status in a clear and engaging speech style. We are particularly interested in non-referential gestures, i.e. gestures not depicting the semantic content of a word (e.g., unfolding arms) and beat gestures, which are claimed to be prosodic in nature (Shattuck-Hufnagel & Ren 2018, Shattuck-Hufnagel et al. 2016). We consider three levels of information status—referential, lexical, and alternative (i.e. contrastive) levels—adopting the RefLex scheme (Riester & Baumann 2017). We address the following research questions:

- (1) What is the relation between co-speech gestures and accent types? Are gestures more likely to occur with H* or L+H* than L* or !H*?
- (2) How do co-speech gestures encode information status? Are they more likely to occur with newer information than with more given information?

2. Methods. This section consists of five parts. The speech material is described in Chapter 2.1. The annotation schemes of co-speech gestures, pitch accents, and information status are introduced in Chapter 2.2., 2.3., and 2.4., respectively. The data analysis is presented in Chapter 2.5.

2.1. SPEECH MATERIAL. A speech entitled “Try something new for thirty days” was obtained from the TED Talks web page (https://www.ted.com/talks/matt_cutts_try_something_new_for_30_days). The speech was delivered by a male speaker of American English in a clear and engaging manner (361 words, $t = 2' 25''$). It was assumed to be rehearsed and assisted by a teleprompter. TED talks were considered an ideal testbed to examine information status, pitch accents, and co-speech gestures because speakers would deliver their speeches in a coherent and efficient manner for an audience without shared background knowledge.

2.2. ANNOTATION OF GESTURES. Gestures were labeled using the Illinois Gesture Annotation Scheme¹, which is a function-oriented annotation scheme. The scheme allows us to annotate six types of hand gestures: iconic, deictic, emblematic, self-touching, objective, and beat gestures. ‘Objective gesture’ describes a gesture that is not depicting semantic content of speech. ‘Beat gesture’ indicates a gesture with rhythmic phases, which can be added to other types of gesture. In this TED talk, four types of hand gesture, i.e. iconic, deictic, objective, and beat gestures, were observed. Two of these gesture types, namely objective and beat gestures ($n = 57$), were considered non-referential and were analyzed in this study. Referential gestures ($n = 4$) and the sections where the speaker was invisible ($n = 138$) were excluded from the analysis.

2.3. ANNOTATION OF PROMINENCE. Pitch accents were labeled by two linguistic experts following the ToBI Annotation Conventions (Veilleux et al. 2006). Six accent types, H*, L*, !H*, L+H*, H+!H*, and L*+H, were observed in this speech. Due to the small sample size, two accent types, H+!H* ($n = 1$) and L*+H ($n = 1$), were reassigned to other accent types: H+!H*

¹ We appreciate Jennifer Cole allowing us to use the Illinois Annotation Scheme. The scheme was developed for the Illinois Games Corpus, which was supported by Volkswagen Stiftung (see PAGE project: http://page.home.amu.edu.pl/?page_id=52).

to !H*, since it displays the same starred tone, and L*+H to the group of L+H* accents, due to the similar contour shape.

2.4. ANNOTATION OF INFORMATION STATUS. The information status of a word in the entire TED talk was labeled based on a simplified version of the RefLex Scheme (Riester & Baumann 2017). This scheme enables us to describe two levels of discourse meaning, namely layers of referential and lexical givenness. The ‘alternative’ or contrastive level was added for completeness. The referential, lexical, and alternative levels consist of different numbers of labels. Table 2 shows the descriptions of the labels as well as examples (see Riester & Baumann 2017, Rooth 1992). The labels are displayed by increasing newness for each level (i.e., *r-given* < *r-bridging* < *r-unused* < *r-new* for the referential level, *l-given* < *l-new* for the lexical level).

| Level | Label | Description | Example |
|-------------|------------|--|--|
| Referential | R-given | Referent present in the prior context | I met a man yesterday. The man told me a story. |
| | R-bridging | Referent accessible from the prior context | The referee lost control over the football match . |
| | R-unused | Unique and new referent in context (definite expression) | the highest mountain of the Himalayan |
| | R-new | Non-unique and new referent in context (indefinite expression) | There was police in front of the building. |
| Lexical | L-given | Active expression in context | Look at the funny dog over there! I like that dog . |
| | L-new | Non-active expression in context | I waled into my hotel room. The chandeliers sparked brightly. |
| Alternative | Alt | Alternative referent present in context | An American farmer was talking to a Candian farmer . |

Table 2: Annotation labels of information status

2.5. ANALYSIS. In order to examine whether the number of occurrences of co-speech gestures significantly differs across accent types and information status, five Fisher’s exact tests were submitted in R (R Core Team 2019). One test for the relation between gestures and accent types was performed based on Table 3. Three tests for the relation between gestures and information status were run separately for the referential, lexical, and alternative levels based on Table 4. Three new labels, *r-none*, *l-none*, and *alt-none*, were added to capture the words not conveying referential, lexical, or contrastive information, respectively. For further analysis, another Fisher’s exact test was carried out based on Table 5 to examine the occurrence of gestures in the presence (or absence) of pitch accents for delivering information status. *IS* indicates the words conveying referential, lexical, or contrastive information in this speech. *IS-none* includes the words not delivering any information status.

| Accent type | Gesture absent | Gesture present |
|-------------|----------------|-----------------|
| Unaccented | 104 | 4 |
| L* | 8 | 4 |
| !H* | 15 | 5 |
| H* | 20 | 21 |
| L+H* | 16 | 23 |

Table 3. Occurrences of gestures as a function of accent types

| Level | Label | Gesture absent | Gesture present |
|-------------|------------|----------------|-----------------|
| Referential | R-none | 101 | 23 |
| | R-given | 32 | 8 |
| | R-bridging | 2 | 2 |
| | R-unused | 9 | 10 |
| | R-new | 19 | 14 |
| Lexical | L-none | 111 | 16 |
| | L-given | 13 | 7 |
| | L-new | 39 | 34 |
| Alternative | Alt-none | 158 | 47 |
| | Alt | 5 | 10 |

Table 4. Occurrences of gestures as a function of information status

| Label | Accent & gesture absent | Accent present only | Gesture present only | Accent & gesture present |
|---------|-------------------------|---------------------|----------------------|--------------------------|
| IS-none | 65 | 8 | 2 | 8 |
| IS | 39 | 51 | 2 | 45 |

Table 5. Occurrences of gestures and accents as a function of information status

3. Results. This section is composed of three parts. We first present the relation between co-speech gestures and accent types in Chapter 3.1. and then the relation between gestures and three levels of information status in Chapter 3.2. The relation between communication modalities and information status is discussed in Chapter 3.3.

3.1. CO-SPEECH GESTURES AND ACCENT TYPES. It was expected that the occurrence of gestures would increase along the Prominence hierarchy. Put differently, gestures would occur with H* and L+H* more frequently than with L* and !H*. Figure 1 shows the relation between gestures and accent types in the TED talk. The absence and presence of gestures are shown in blue and red bars, respectively. On the y-axis, accent types are arranged in increasing order of perceived prominence (from top to bottom) as in the Prominence hierarchy. On the x-axis, the gesture occurrence for accent types is displayed in percent. The percentage was obtained from the division of the number of words occurring with (or without) gestures by the entire number of words occurring with pitch accents. The calculation was based on Table 3 in Chapter 2.5. Results supported our prediction. The Fisher's exact test returned significant differences in the occurrence of gestures (i.e., absence versus presence of gestures) for the accent types ($p < .01$). In Figure 1, we observe that the proportion of gestures increases along the Prominence hierarchy.

Co-speech gestures were more likely to be accompanied by L+H* than H*, which in turn was more frequent than !H*. They were the least likely to occur with unaccented words. The gesture occurrence was unexpectedly more frequent for L* than for !H*. In Bishop et al. (in press) and Hualde et al. (2016), L* was found to be perceived as prominent when it was located in a structurally strong position (i.e., nuclear pitch accent) for yes-no questions or continuation (L* H-H%), presumably since it sounds like a rising (L*+H) pitch accent (and not like a low one). Accordingly, in our TED talk, L* was frequently accompanied by gestures if it was followed by a continuation pattern, i.e. a high phrase boundary. Also, we observe that the relation between gestures and prominence is probabilistic, not exclusive. Although L+H* is highest in the Prominence hierarchy, only 59% of the accents co-occurred with gestures in this speech. Unaccented words, ranked lowest in the Prominence hierarchy, were still accompanied by gestures in this TED talk (4%). These unaccented words occurring with gestures were found to be preceded by accented words with the same type of gestures.

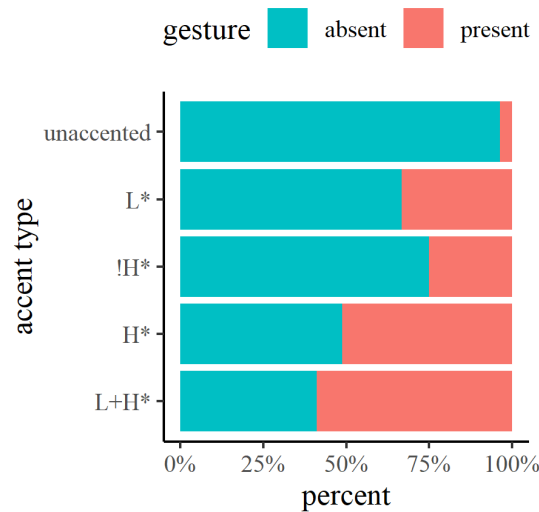


Figure 1. Relation between gestures and accent types

3.2. CO-SPEECH GESTURES AND INFORMATION STATUS. It was predicted that the occurrence of gestures would increase along the Information status hierarchy. In other words, gestures would occur with newer information more frequently than with given information. Figures 2-4 show the relation between gesture occurrence and information status at the referential, lexical, and alternative levels, respectively. Across Figures 2-4, the x-axis indicates the percentage of gestures conveying information status based on Table 4 in Chapter 2.5. On the y-axis, the IS labels are displayed in increasing order of IS, as in the Information status hierarchy. The bars indicate the absence of gestures in blue and the presence of gestures in red. Our prediction was supported. Three Fisher's exact tests show significant differences in gesture occurrence for information status ($p < .01$). Figure 2 indicates that gestures are more likely to occur with words delivering referentially accessible (*r-bridging*) or referentially new information (*r-unused* and *r-new*) than words with referentially given information (*r-given*) or words not delivering referential information status (*r-none*). In the TED talk, the occurrence of gestures was unexpectedly more frequent for *r-bridging* (50%) and *r-unused* (53%) than for *r-new* (42%). Recall that *r-bridging* describes accessible words from the prior context (e.g., *a bus – the driver*). *R-unused* captures new and unique entities (e.g., a commonly known place like *Rome*), which correspond to proper

nouns. Hirschberg (1993) found in a corpus of broadcast news speech that proper nouns tend to be produced with prosodic prominence to draw the interlocutor’s attention to the topic being discussed. Similarly, the expressions for giving explanations or introducing persons or places could be accompanied by gestures to attract attention from the large audience in the TED talk.

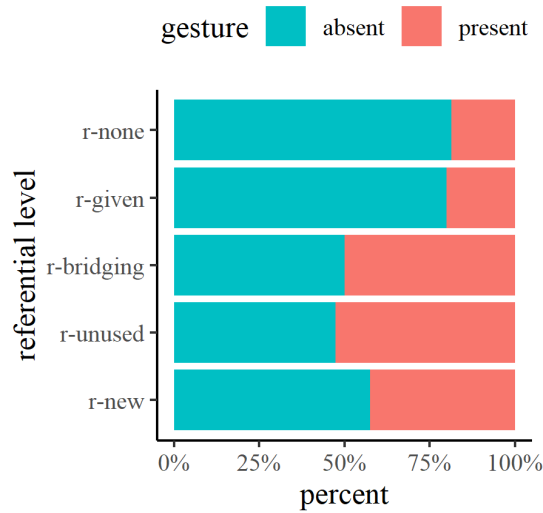


Figure 2. Relation between gestures and referential information status

In Figure 3, we can see that gestures are more likely to be accompanied by lexically new information (*l-new*) than by lexically given information (*l-given*). They were the least likely to occur with words not carrying lexical information (*l-none*). There was a stepwise increase of gesture occurrence across *l-none*, *l-given*, and *l-new*. This differed from the binary patterns in Figure 2, showing that gestures tended to occur with accessible and new information (*r-bridging*, *r-unused*, and *r-new*) more frequently than with given and no information (*r-none* and *r-given*). In Figures 2 and 3, the gesture occurrence does not exceed 50% for both referentially new (42%) and lexically new (47%) information although new information occupies the highest rank in the Information status hierarchy.

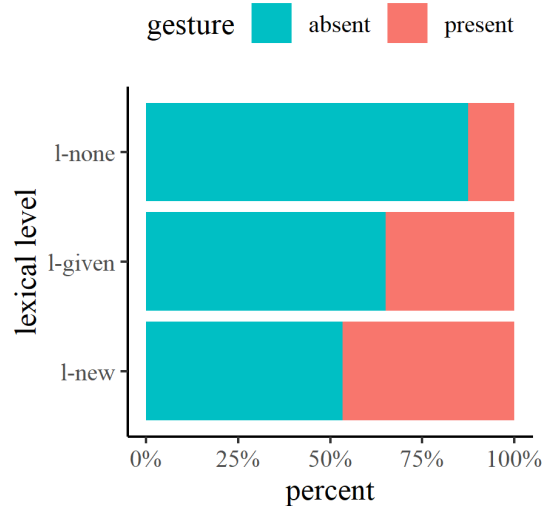


Figure 3. Relation between gestures and lexical information status

Figure 4 shows that the occurrence of gestures is higher for words conveying contrastive information status (*alt*, 67%) than words which do not (*alt-none*, 23%). Across all the IS labels in Figures 2-4, *alt* is the most frequently associated with co-speech gestures (67%). In other words, the speaker was most likely to use a gesture if he was to deliver some contrastive meaning. In Figures 2-4, we observe again that the relation between co-speech gestures and information status is not one-to-one. Despite the relatively strong association, only 67% of contrastive words were delivered in the presence of gestures.

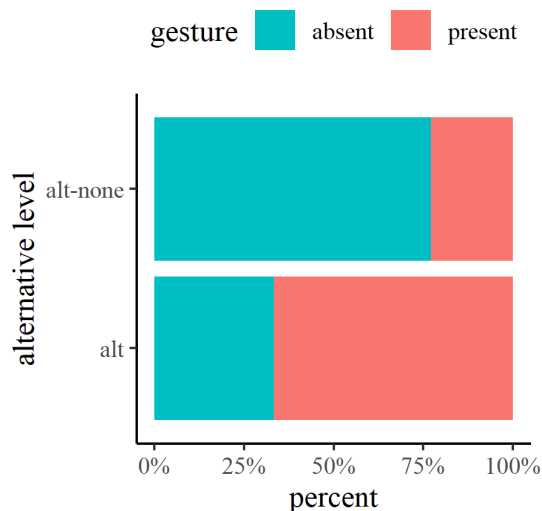


Figure 4. Relation between gestures and contrastive information status

3.3. COMMUNICATION MODALITIES AND INFORMATION STATUS. To complete our analysis, we further examined whether or not gestures would occur with pitch accents to convey information status. In Figure 5, the x-axis indicates the proportions based on Table 5 in Section 2.5. On the y-axis, as mentioned above, *IS* stands for the words delivering referential, lexical, or alternative information status and *IS-none* for the words not conveying any IS. Color-coded bars represent different communication modalities: purple bars for the words that did not occur with accents and gestures, blue bars for the words that occurred with accents only, green bars for the words with gestures only, and red bars for the words with both accents and gestures. The Fisher's exact test shows that the distribution of communication modalities significantly differs in information status ($p < .01$). As Figure 5 indicates, information status (*IS*) is most frequently conveyed by accents only (37%), followed by the accents that co-occur with gestures (33%). Information status was rarely marked by gestures only (1%).

4. Discussion. In this study, we examined how the occurrence of co-speech gestures varied with accent types and information status in a clear and engaging speech style. We first expected that gestures would occur more frequently with H^* and $L+H^*$ than with L^* and $!H^*$. Our results showed that there was an increasing tendency of gesture occurrence as perceived prominence increased ($L^* < !H^* < H^* < L+H^*$). This might not be surprising considering the previous studies claiming that the two modes, co-speech gestures and prosodic prominence, stem from the same source, namely a specific intention of a speaker, often with the aim to attract some attention (de Ruiter 2000, Iverson & Thelen 1999, Loehr 2012, McClave 1991, McNeill 1992, among many others). However, we found (see Figure 5) that information status was most frequently marked

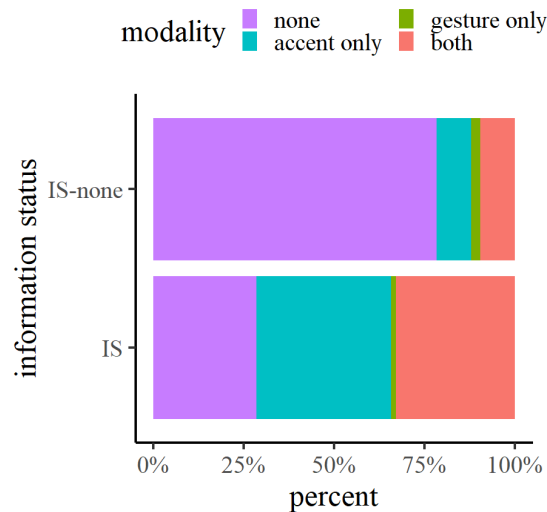


Figure 5. Relation between communication modalities and information status

by pitch accents only (51 words), followed by pitch accents that co-occurred with gestures (45 words). This suggests that among all the accented words (96 words), only 45% of accented words were accompanied by gestures. If co-speech gestures and pitch accents simply were two different modes of expressing a speaker's intention, all the accented words would be accompanied by gestures, which did not turn out to be the case. The speaker in this TED talk seemed to rely on pitch accents to a greater extent than on gestures to encode information status. What is the reason why some accented words conveying information status are not accompanied by gestures? Our examination revealed that the accent types tended to differ between a) accented words accompanied by accents only and b) accented words accompanied by both accents and gestures. Lower ranked accent types in the Prominence hierarchy, L^* and $!H^*$, were clearly less frequent for the words with both accent and gestures (8 words) than the words with accents only (21 words). Higher ranked accent types, H^* and $L+H^*$, were slightly more frequent for the words with both accents and gestures (37 words) than for the words with accents only (30 words). This suggests that we need to consider not only the presence (or absence) of pitch accents but also the *type* of pitch accent to predict the occurrence of gestures in future research.

Our findings also confirm the probabilistic relation between gestures and accent types. In this speech, we found that gestures occurred with all the accent types. Only the probability of gesture occurrence differed across the types of accent. The likelihood of gesture occurrence increased as the strength of perceived prominence increased, as discussed above. The probabilistic association between co-speech gestures and accent types can be due to the complex nature of prosodic prominence. The assignment of prosodic prominence reflects the semantic or pragmatic meaning of a word in discourse context, but it can also interact with other linguistic and non-linguistic factors including speech rhythm (Büring 2007, Calhoun 2006, Sityaev 2000), parts-of-speech (Hirschberg 1993, Roy et al. 2017, Sityaev 2000), speech styles (de Ruiter 2015, Hirschberg 1993, Im et al. 2018), and speaker's emotion (Chodroff & Cole 2018). Further examination showed that the accented words not delivering information status were accompanied by gestures half of the time (50%). In other words, gestures can occur with accents to deliver linguistic and non-linguistic meanings other than information status, which requires more

comprehensive research in the future. Another possible reason for the probabilistic relation between the two modes can be the differences between prosodic and gestural systems. Consider the findings shown in Figure 1 that there were only a few occasions when unaccented words occurred with gestures (4%). These words were found to be preceded by accented words accompanied by the same type of gesture. Put differently, in the two adjacent words, the speaker in this speech assigned prominence alternatively (to the preceding words only), while allocating gestures repeatedly (to both words). It can be inherently easier for a gesture to be carried over from a word into the following word, creating its own rhythmic phrases (e.g., beats). This differs from the prosodic system, where the assignment of prominence in two adjacent words tends to be avoided (Shattuck-Hufnagel et al. 1994, Nespor & Vogel 1989).

Finally, our results show that the occurrence of gestures varied with the information status of a word in discourse context. It was expected to find that gestures would more frequently co-occur with new or contrastive information than with given or accessible information. As to referential information status, however, there was no increase of the number of gestures with increasing newness of the expressions. Gestures were more likely to occur with *unused* and *bridging* items than with *new* ones. As to the level of lexical information status, gesture occurrence increased with increasing newness. Thus, the patterns of gesture occurrence differed between the referential and lexical levels, which can be considered as supporting evidence for the RefLex Scheme (Riester & Baumann 2017). Bergmann and Kopp (2006), using given-new binary distinctions, found systematic patterns of gesture occurrence for entities only, but not for the other expressions encoding action, property, amount, etc. The present study broadens our understanding of gesture occurrence (1) between referential and lexical information status and (2) among given, accessible, and new information, adopting a more elaborated annotation scheme of information status. As to the level of alternative information status, contrastive words were frequently associated with gestures (67%). This was the most frequent case of gesture occurrence, followed by *r-unused* (53%), *r-bridging* (50%), *l-new* (47%), and *r-new* (42%). Overall, our results suggest that words whose information status has to be derived from previous discourse, which introduce new and unique entities, and which mark contrast are strong candidates for predicting the occurrence of gestures. Loehr (2012) observed that contrastive or emphatic words marked by H* or L+H* were associated with beats in a corpus of spontaneous speech style. Indeed, in this clear speech style, contrastive words marked by H* or L+H* were very frequently associated with gestures including beats. However, there was still a case where a contrastive word marked by H* or L+H* was not accompanied by gestures. This again suggests the probabilistic relation between gestures, accents, and information status.

5. Conclusion. In this study, we investigated the occurrence of co-speech gestures as a function of accent types and information status in a clear and engaging speech style. As to information status, we considered the referential, lexical, and contrastive levels of activation for each word in the corpus. Our results show that gestures generally co-occur with pitch accents in conveying information status, however more frequently with those accent types that are more likely to be perceived as prominent ($L^* < !H^* < H^*$ and $L+H^*$). Moreover, in the TED talk data, gestures tended to occur with new, accessible or contrastive information more frequently than with given information. The patterns of gesture occurrence differed between the referential and lexical level of information status, though.

Overall, this study suggests (1) that co-speech gestures contribute to the probabilistic encoding of a word's information status in conjunction with pitch accents, and (2) that more

complex distinctions of information status need to be considered in future research. In any case, the classification of co-speech gestures as “prosodic” as used by Shattuck-Hufnagel & Ren (2018) appears to be appropriate.

References

- Baumann, Stefan & Arndt Riester. 2013. Coreference, lexical givenness and prosody in German. *Lingua* 136. 16–37.
- Bergmann, Kirsten & Stefan Kopp. 2006. Verbal or visual: How information is distributed across speech and gesture in spatial dialog. *Proceedings of the 10th Workshop on the Semantics and Pragmatics of Dialogue*. 90–97.
- Bishop, Jason, Grace Kuo & Boram Kim. In press. Phonology, phonetics, and signal-extrinsic factors in the perception of prosodic prominence: Evidence from Rapid Prosody Transcription. *Journal of Phonetics*.
- Büring, Daniel. 2007. Intonation, semantics and information structure. In Gillian Ramchand & Charles Reiss (eds.), *The Oxford handbook of linguistic interfaces*. 445–474. Oxford, UK: Oxford University Press.
- Calhoun, Sasha. 2006. *Information structure and the prosodic structure of English: A probabilistic relationship*. Edinburgh, UK: University of Edinburgh dissertation.
- Cangemi, Francesco & Martine Grice. 2016. The importance of a distributional approach to categoriality in Autosegmental-Metrical accounts of intonation. *Laboratory Phonology: Journal of the Association for Laboratory Phonology* 7(1). 9. <https://doi.org/10.5334/labphon.28>.
- Chafe, Wallace. 1976. Givenness, contrastiveness, definiteness, subjects, topics, and point of view. In Charles N. Li (ed.), *Subject and topic*. 25–55. New York, NY: Academic Press.
- Chafe, Wallace. 1994. *Discourse, consciousness, and time*. Chicago, IL/London, UK: University of Chicago Press.
- Chodroff, Eleanor R. & Jennifer Cole. 2018. Information structure, affect, and prenuclear prominence in American English. *Proceedings of Interspeech 2018*. 1848–1852. <https://doi.org/10.21437/Interspeech.2018-1529>.
- de Ruiter, Jan Peter. 2000. The production of gesture and speech. In David McNeill (ed.), *Language and gesture*. 248–311. Cambridge, UK: Cambridge University Press
- de Ruiter, Jan Peter, Matthijs L. Noordzij, Sarah Newman-Norlund, Roger Newman-Norlund, Peter Hagoort, Stephen C. Levinson & Ivan Toni. 2010. Exploring the cognitive infrastructure of communication. *Interaction Studies* 11(1). 51–77. <https://doi.org/10.1075/is.11.1.05rui>.
- de Ruiter, Laura E. 2015. Information status marking in spontaneous vs. read speech in storytelling tasks—Evidence from intonation analysis using GToBI. *Journal of Phonetics* 48. 29–44. <https://doi.org/10.1016/j.wocn.2014.10.008>.
- Halliday, Michael A. K. & Ruqaiya Hasan. 1976. *Cohesion in English*. London: Longman.
- Halliday, Michael A. K. & Christian M. I. M. Matthiessen. 2004. *An introduction to functional grammar*. London: Edward Arnold.
- Hirschberg, Julia. 1993. Pitch accent in context predicting intonational prominence from text. *Artificial Intelligence* 63(1-2). 305–340. [https://doi.org/10.1016/0004-3702\(93\)90020-C](https://doi.org/10.1016/0004-3702(93)90020-C).
- Holler, Judith & Rachel Stevens. 2007. The effect of common ground on how speakers use gesture. *Journal of Language and Social Psychology* 26. 4–27. <https://doi.org/10.1177/0261927X06296428>.

- Hualde, José. I., Jennifer Cole, Caroline L. Smith, Christopher D. Eager, Timothy Mahrt & Ricardo Napoleão de Souza. 2016. The perception of phrasal prominence in English, Spanish and French conversational speech. *Proceedings of Speech Prosody* 8. 459–463. <https://doi.org/10.21437/SpeechProsody.2016-94>.
- Im, Suyeon, Jennifer Cole & Stefan Baumann. 2018. The probabilistic relationship between pitch accents and information status in public speech. *Proceedings of Speech Prosody* 9. 508–511. <https://doi.org/10.21437/SpeechProsody.2018-103>.
- Iverson, Jana M. & Esther Thelen. 1999. Hand, mouth and brain—the dynamic emergence of speech and gesture. *Journal of Consciousness Studies* 6(11-12). 19–40.
- Jannedy, Stefanie & Norma Mendoza-Denton. 2005. Structuring information through gesture and intonation. *Interdisciplinary Studies on Information Structure (ISIS)* 3. 199–244.
- Kendon, Adam. 2004. *Gesture—visible action as utterance*. Cambridge, UK: Cambridge University Press.
- Lambrecht, Knud. 1994. *Information structure and sentence form: Topic, focus, and the mental representations of discourse referents*. Cambridge, UK: Cambridge University Press.
- Loehr, Daniel P. 2004. *Gesture and intonation*. Washington, D.C.: Georgetown University dissertation.
- Loehr, Daniel P. 2012. Temporal, structural, and pragmatic synchrony between intonation and gesture. *Laboratory Phonology: Journal of the Association for Laboratory Phonology* 3. 71–89.
- McClave, Evelyn. 1991. *Intonation and gesture*. Washington, D.C.: Georgetown University dissertation.
- McClave, Evelyn. 1994. Gestural beats: The rhythm hypothesis. *Journal of Psycholinguistic Research* 23(1). 45–66. <https://doi.org/10.1007/BF02143175>.
- McNeill, David. 1992. *Hand and mind: What gestures reveal about thought*. Chicago, IL: University of Chicago Press.
- Nespor, Marina & Irene Vogel. 1989. On clashes and lapses. *Phonology* 6(1). 69–116. <https://doi.org/10.1017/S0952675700000956>.
- Pierrehumbert, Janet & Julia B. Hirschberg. 1990. The meaning of intonational contours in the interpretation of discourse. In Philip R. Cohen, Jerry Morgan & Martha E. Pollack (eds.), *Intentions in Communication*. 271–311. Cambridge, MA: MIT Press.
- Prince, Ellen F. 1981. Towards a taxonomy of given-new information. In Peter Cole (ed.), *Radical pragmatics*. 223–256. New York, NY: Academic Press.
- R Core Team. 2019. R: A language and environment for statistical computing. <https://www.R-project.org>. (23 February, 2020).
- Riester, Arndt & Stefan Baumann. 2017. *The RefLex Scheme—annotation guidelines*. University of Stuttgart.
- Rooth, Mats. 1992. A theory of focus interpretation. *Natural Language Semantics* 1(1). 75–116. <https://doi.org/10.1007/BF02342617>.
- Roy, Joseph, Jennifer Cole & Tim Mahrt. 2017. Individual differences and patterns of convergence in prosody perception. *Laboratory Phonology: Journal of the Association for Laboratory Phonology* 8. 1–36. <http://doi.org/10.5334/labphon.108>.
- Schwarzschild, Roger. 1999. GIVENness, AvoidF, and other constraints on the placement of accent. *Natural Language Semantics* 7(2). 141–177. <https://doi.org/10.1023/A:1008370902407>.

- Shattuck-Hufnagel, Stefanie, Mari Ostendorf & Ken Ross. 1994. Stress shift and early pitch accent placement in lexical items in American English. *Journal of Phonetics* 22(4). 357–388. [https://doi.org/10.1016/S0095-4470\(19\)30291-8](https://doi.org/10.1016/S0095-4470(19)30291-8).
- Shattuck-Hufnagel, Stefanie & Ada Ren. 2018. The prosodic characteristics of non-referential co-speech gestures in a sample of academic-lecture-style speech. *Frontiers in Psychology* 9. 1514. <https://doi.org/10.3389/fpsyg.2018.01514>.
- Shattuck-Hufnagel, Stefanie, Ada Ren, Mili Mathew, Ivan Yuen & Katherine Demuth. 2016. Non-referential gestures in adult and child speech: Are they prosodic? In Jon Barnes, Alejna Brugos, Stefanie Shattuck-Hufnagel & Nanette Veilleux (eds.), *Speech Prosody 2016*. 836–839. <https://doi.org/10.21437/SpeechProsody.2016-171>.
- Shattuck-Hufnagel, Stefanie, Yelena Yasinnik, Nanette Veilleux & Margaret Renwick. 2007. A method for studying the time alignment of gestures and prosody in American English: ‘hits’ and pitch accents in academic-lecture-style speech. In Anna Esposito, Maja Bratanić, Eric Keller & Maria Marinaro (eds.), *Fundamentals of verbal and nonverbal communication and the biometric issue*. In *NATO security through science series E: Human and societal dynamics* 18. Washington, D.C.: IOS Press.
- Sityaev, Dmitry. 2000. The relationship between accentuation and information status of discourse referents: A corpus-based study. *UCL Working Papers in Linguistics* 12. 285–304.
- Veilleux, Nanette, Stefanie Shattuck-Hufnagel & Alejna Brugos. 2006. 6.911 Transcribing prosodic structure of spoken utterances with ToBI. <https://ocw.mit.edu>. (23 February, 2020).
- Yassinik, Yelena, Margaret Renwick & Stefanie Shattuck-Hufnagel. 2004. The timing of speech-accompanying gesture with respect to prosody. *Proceedings of the International Conference: From Sound to Sense*. 50. 10–13.