Memory & Cognition (2019) 47:1515–1530
https://doi.org/10.3758/s13421-019-00945-1

Listeners consider alternative speaker productions in discourse comprehension and memory: Evidence from beat gesture and pitch accenting

Laura M. Morett[1] · Scott H. Fraundorf[2]

Published online: 18 June 2019

Abstract
Cues to emphasis, such as beat gesture and contrastive pitch accenting, play an important role in constraining what comprehenders remember from a discourse. One possibility is that these cues are used in a purely bottom-up manner in which additional attention is devoted to emphasized material. Another possibility is that comprehenders use top-down expectations of what cues might be expected in the current communicative context, such that the absence of an expected cue may serve as an indicator that material is unimportant. We independently manipulated two cues conveying emphasis – beat gesture and contrastive pitch accenting – to examine how they affected memory for information in a discourse. When beat gesture was present in some cases (Experiment 1), contrastive pitch accenting facilitated memory when beat gesture was present but not when beat gesture was absent. By contrast, when beat gesture was never present (Experiment 2), contrastive pitch accenting facilitated memory even though stimuli were identical to those in which beat gesture was absent in Experiment 1. Together, these results indicate that which cues *could* be produced affects interpretation even when these cues are absent, indicating that top-down expectations influence cue integration, consistent with emerging data-explanation views of language processing.

Keywords Psycholinguistics · Discourse processing · Embodied cognition · Language comprehension

Introduction

A fundamental challenge of human language is that it is incredibly rich and complex, but that it is impossible to attend to and remember everything that it communicates. Fortunately, language contains many cues that may provide information about what should be attended to and remembered for later on (Givón, 1992). For instance, prior work has established that listeners' memory for discourse is guided both by contrastive pitch accents in speech and by beat gestures produced with the hands (e.g., Feyereisen, 2006; Fraundorf, Watson, & Benjamin, 2010, 2012). While there is evidence that these cues contribute individually to memory, discourses often contain multiple types of such cues, and it is unclear how they are integrated. One possibility is these cues function in a purely bottom-up manner, attracting attention to particular parts of the discourse when they are present (Givón, 1992). In contrast, another possibility, suggested by emerging data-explanation views of language processing (Farmer, Brown, & Tanenhaus, 2013), is that top-down expectations about what kinds of cues a talker typically produces may render even the *absence* of a cue informative, such that the absence of one typically produced cue (e.g., beat gesture) might affect how other cues (e.g., pitch accenting) are interpreted.

In the present study, we examine how pitch accents and beat gesture are integrated and how they affect memory for discourse, with an eye towards what these effects indicate about language processing more broadly. We introduce a multimodal version of a paradigm previously used to study the effect of pitch accenting on memory for discourse (Fraundorf et al., 2010), so that we can examine the combined effects of contrastive pitch accents and beat gesture using carefully controlled experimental stimuli. In the remainder of the *Introduction*, we discuss what is already known about pitch accenting and beat gesture individually before turning to the question of how these cues might be integrated.

✉ Laura M. Morett
lmorett@ua.edu

[1] Department of Educational Studies in Psychology, Research Methodology, and Counseling, University of Alabama, Tuscaloosa, AL 35401, USA

[2] Learning Research and Development Center, Department of Psychology, University of Pittsburgh, Pittsburgh, PA, USA

Pitch accenting and its effect on discourse processing

One cue that speakers often use to indicate the discourse status of referents in spoken English is a *pitch accent*, a phonological construct realized acoustically via changes in fundamental frequency, duration, and intensity (for review, see Ladd, 2008). Many theories distinguish between different types of pitch accents. For example, the ToBI framework for intonational transcription of English distinguishes an H* accent, which consists of a high pitch target with fundamental frequency (F_0) high in the speaker's range, from a L+H* accent, which consists of an initial low pitch (the L) followed by a sharp rise to a high target on the accented syllable (the H*).[1] It has been proposed that the H* accent is associated with information that is new to a discourse and not contrastive, whereas the L+H* accent is associated with information that specifically contrasts with something else in the discourse (Brown, 1983; Pierrehumbert & Hirschberg, 1990). For example, consider the discourse below:

(1a) [S1] What did Marjorie have for lunch?
(1b) [S2] She had a salad.
(2a) [S1] Did you say she had a sandwich?
(2b) [S2] No, I said she had a SALAD.

In (1b), *salad* is new information and would likely carry a presentational H* accent. In (2b), however, *salad* contrasts with a referent previously mentioned in (2a), *sandwich,* and would likely carry a contrastive L+H* accent.

Evidence suggests that speakers produce contrastive accents deliberately to signal such contrastive information to their addressee (Kaland, Krahmer, & Swerts, 2011), and that listeners are indeed sensitive to this distinction (for a review, see Gotzner & Spalek, 2019). In online comprehension, contrastive accenting directs listeners' attention to referents that contrast with a specific previous referent, whereas presentational accents direct attention to new referents more broadly, as revealed by eye tracking in the visual world (Ito, Jincho, Minai, Yamane, & Mazuka, 2012; Ito & Speer, 2008; Watson, Tanenhaus, & Gunlogson, 2008; Weber, Braun, & Crocker, 2006) and cross-modal priming (Braun & Tagliapietra, 2010; Husband & Ferreira, 2016). Similarly, in offline representations of discourse, contrastively accented referents are remembered better than presentationally-accented referents, particularly when the mnemonic task requires ruling out a salient contrast item (Fraundorf et al., 2010, 2012; Lee & Fraundorf, 2016; Lee & Snedeker, 2016; Sanford, Sanford, Molle, & Emmott, 2006). For instance, in a particularly relevant study, Fraundorf et al. (2010) tested memory for spoken discourses. Each discourse consisted of a context passage, such as (3) below, introducing two pairs of alternatives (*British* and *French; Malaysia* and *Indonesia*), and a continuation sentence, such as (4) below, in which one alternative from each pair is mentioned.

(3) Both the British and the French biologists had been searching Malaysia and Indonesia for the endangered monkeys.
(4) Finally, the BRITISH spotted one of the monkeys in MALAYSIA and planted a radio tag on it.

Critically, Fraundorf et al. (2010) manipulated whether each of the alternatives mentioned in the continuation sentence carried contrastive versus presentational pitch accenting in order to examine their effects on subsequent memory. After hearing all of the discourses, participants completed a recognition memory test in which they saw each discourse represented as text with each critical word in the continuation passage replaced with a blank, such as in (5). Participants completed a two-alternative forced-choice task to identify each critical word from each discourse.

(5) Both the British and the French biologists had been searching Malaysia and Indonesia for the endangered monkeys. Finally, the (British/French) spotted one of the monkeys in _____ (Malaysia/Indonesia) and planted a radio tag on it.

Recognition memory for critical words was better when they received contrastive accenting than presentational accenting in the discourse, regardless of their position and the pitch accenting of the other word. A further experiment (Fraundorf et al., 2010; Experiment 3) revealed that contrastive accenting enhanced memory specifically by facilitating rejections of the salient alternative (e.g., *French* in the example above), leading the authors to conclude that contrastive accenting contributes to discourse memory because it prompts comprehenders to represent something about an important salient alternative in the discourse (i.e., remembering that it was *not* the French scientists who found the monkey; see also (Spalek, Gotzner, & Wartenburger, 2014, for similar results with other focusing devices).

Gesture and its effect on discourse processing

Another cue that may be relevant to memory for discourse is beat gesture. According to McNeill's (1992, 2005) widely

[1] Both of these pitch accents can be distinguished from cases in which a word is completely deaccented. Deaccenting is a marked form reserved for information already given in a discourse or otherwise predictable (Schwarzschild, 1999; Steedman, 2000) and would thus be infelicitious in the discourses used in this study.

used taxonomy, *beat gesture* is non-referential gesture that reflects the prosody of accompanying speech. Prototypically, beat gesture takes the form of punctate downward hand flicks, but it can be articulated using other parts of the body (e.g., finger movements, head nods, foot taps), in other orientations (e.g., horizontal, oblique, curved), and with multiple components (Shattuck-Hufnagel & Ren, 2018). In discourse, beat gesture, like pitch accenting, is frequently used to emphasize the importance of select words or phrases, serving as a "gestural yellow highlighter" (McNeill, 2006).

Indeed, there is evidence that beat gesture and pitch accenting in general – as well as contrastive accenting in particular – are closely related in perception and production. While the onset of gesture typically precedes the onset of pitch accenting, the points of maximum extension (*apices*) of gestures are closely temporally aligned with the F_0 peaks of pitch-accented words for beat gesture in particular (Esteve-Gibert & Prieto, 2013; Leonard & Cummins, 2011) as well as for gesture more broadly (Roustan & Dohen, 2010a, 2010b; Rusiewicz, Shaiman, Iverson, & Szuminsky, 2013, 2014). Even 9-month-old infants, who are unable to produce gesture in conjunction with speech, are sensitive to the temporal alignment of deictic (pointing) gesture and syllabic stress (Esteve-Gibert, Prieto, & Pons, 2015). These cues also align at an acoustic level insofar as words with accompanying beat gestures have been found to be produced with higher vowel formants (*F2* and *F3*) and be more likely to be perceived as pitch accented than words without accompanying beat gestures (Krahmer & Swerts, 2007). However, Roustan and Dohen (2010a) found that gesture production has no effect on either articulatory (vocalic target) or acoustic (intensity, duration, F_0) correlates of contrastive pitch accenting. One key difference between these studies may be that beat gesture was always accompanied by pitch accenting in Roustan and Dohen (2010a), whereas beat gesture sometimes occurred without pitch accenting in Krahmer and Swerts (2007). Together, the results of these two studies suggest that comprehenders are sensitive to the frequency with which a particular cue (e.g., pitch accenting) is produced and that the *absence* of one cue (pitch accenting) may influence how another related cue (beat gesture) is interpreted in discourse.

Beyond the temporal alignment of beat gesture and pitch accents, beat gesture also resembles pitch accenting in that it enhances memory for spoken language at the lexical level. For example, beat gesture facilitates discrimination between pairs of L2 words differing minimally in vowel length (Hirata, Kelly, Huang, & Manansala, 2014), and production and observation of spontaneous beat gestures predicts the number of times that novel L2 words are repeated in discourse (Morett, 2014). Moreover, both L1 and L2 words are more likely to be recalled when accompanied by beat gesture or representational gesture (gesture depicting semantic information via form and/or motion) than no gesture (Levantinou & Navarretta, 2016; So et al., 2012). Similarly, within a discourse, children are more likely to recall words accompanied by beat gesture than words unaccompanied by beat gesture (Igualada, Esteve-Gibert, & Prieto, 2017). Considered as a whole, this work suggests that the visual prominence conveyed by beat gesture enhances memory for accompanying spoken words.

However, most of the studies reviewed above concern memory only for individual words. While a number of studies have examined how other types of gesture (e.g., representational gesture) affect understanding of a more complex spoken discourse (e.g., Cohen & Otterbein, 1992; Cook & Tanenhaus, 2009; Feyereisen, 2006; Hostetter & Alibali, 2010; Sueyoshi & Hardison, 2005), only a few studies have examined the effect of beat gesture on higher levels of linguistic representation. These studies have generally indicated that beat gesture does not enhance memory for entire sentences for adults (Feyereisen, 2006), although it may for young children (Vilà-Giménez, Igualada, & Prieto, 2019). However, even if beat gesture does not enhance recognition of entire sentences or discourses at a broad level of meaning, it might facilitate memory for *specific* words or phrases that it accompanies within sentences or discourses. Indeed, recent work using a paradigm similar to that of Fraundorf et al. (2010) provides suggestive evidence: Words from a set of alternatives are more likely to be remembered when they are emphasized with both contrastive pitch accenting and beat gesture than when they are emphasized with contrastive accenting alone or neither cue (Kushch & Prieto, 2016; Llanes-Coromina, Vilà-Giménez, Kushch, Borràs-Comes, & Prieto, 2018). However, because these studies did not vary contrastive accenting and beat gesture orthogonally, it is unclear the extent to which memory benefits reflect individual influences of beat gesture and contrastive accenting or the integration of these cues; we take up that question in the current work.

Integrating cues in memory for discourse

Multiple potential cues to linguistic categories are often available, and relevant values and overall relevance of each of these cues may vary across contexts. For instance, voice onset time (VOT), F_0, and first-formant (F1) onset frequency are all relevant cues to distinguishing voiced consonants from voiceless consonants (see Repp, 1982, for a review), and the prototypical values (Allen, Miller, & DeSteno, 2003) and even the relative importance of these cues (Shultz, Francis, & Llanos, 2012) may differ across talkers. Similarly, with regard to pitch accenting and beat gesture as cues to prominence in a discourse, talkers may vary in *which* cues they produce (e.g., whether speakers use beat gestures to convey prominence) and in *how often* they produce those cues (e.g., some people may produce beat gesture for only the most important parts of a discourse, while others may use it more frequently). Thus, it

is possible that interpretation of any one cue (e.g., pitch accenting) may become more or less important depending on the frequency with which a particular talker uses it or on which other cues are present concurrently or in the talker's repertoire more generally. Here, we orthogonally vary beat gesture and pitch accenting to examine how listeners integrate cues in their memory for a discourse, contrasting several hypotheses derived from more general views of language processing.

One broad view of why devices such as contrastive pitch accenting influence long-term memory for a discourse is that they can be viewed as "processing instructions" (Givón, 1992) about what comprehenders should attend to and remember for the future. This account is inspired by findings such as that of syntactic focus: Words that are syntactically focused (e.g., *It was the* ...) are later remembered better than non-focused words (Birch, Albrecht, & Myers, 2000; Birch & Garnsey, 1995), but see Almor and Eimas (2008) for a counterexample. Critically, this hypothesis posits that the presence of particular cues is critical because these devices function through their salience or ability to capture attention; for instance, when listeners encounter a beat gesture, their attention to and memory for the accompanying speech stream increases. This hypothesis permits that multiple cues, when present, might be combined in several different ways: They may have additive effects (whereby memory is enhanced by the combined effect of each cue acting individually), a superadditive effect (whereby memory is enhanced beyond the combined amounts of all cues), or a subadditive effect (whereby each cue is enhanced less than the combined amounts of all cues). In all cases, however, listeners' cue interpretation should be based on the occurrence of cues themselves (a bottom-up effect) rather than inferences about the circumstances under which certain cues are produced (a top-down effect).

By contrast, emerging data-explanation views of language processing (Farmer et al., 2013), and of cognition more broadly (Clark, 2013), propose that the goal of language comprehension is to explain the linguistic input (or data) by modeling the underlying communicative intent. In this view, cues such as beat gesture are relevant to discourse processing and memory because they reflect the talker's communicative intent; for example, that a particular point is especially important to the talker. Thus, the same cue (e.g., beat gesture) could vary in its relevance for later memory depending on what information about the talker's intent is supported by the context. This data-explanation view predicts that even the absence of a particular cue may be informative if it conveys information about the talker's communicative intent; namely, if the talker would have produced the cue had the material been important. For example, if your coworker usually lowers her voice when she's having a confidential conversation, her failure to lower her voice indicates that the conversation she's engaged in is probably not confidential.

One way to formalize this intuition (though not one required in particular by our work) is Bayes's theorem (Rohde & Kurumada, 2018). Suppose a talker produces a proposition without emphasizing it with a beat gesture. How probable is it that this information is particularly important? Bayes's rule (example 6, below) gives the optimal inference as proportional to the overall probability that information in a spoken discourse is important, *P*(*Important*), times the probability that no gesture is produced if the speaker regards a proposition as important, *P*(*No Gesture* | *Important*). If the talker never uses gesture for emphasis, *P(No Gesture* | *Important*) is 1, and the right-hand side of the equation reduces to P (*Important*), the baseline rate at which information in spoken discourse is important (corresponding to the top panel of Fig. 1). However, if the talker emphasizes spoken information with beat gesture half the time (as in our Experiment 1), *P*(*No Gesture* | *Important*) is 0.5, and the right-hand side of the equation reduces to *less* than the baseline rate of importance (corresponding to the bottom panel of Fig. 1). That is, the absence of an expected beat gesture can serve as a cue that the information is less likely to be important than it otherwise would be.

$$(6)\ P(Important \mid No\ Gesture) \propto P(Important) * P(No\ Gesture \mid Important)$$

Supporting this view, the availability of alternative but unspoken utterances has been argued to explain interpretation of vowel shifts in regional accents (Trude & Brown-Schmidt, 2012), pragmatic processing (Bergen, Goodman, & Levy, 2012), and scalar implicature interpretation (Degen & Tanenhaus, 2016). These effects extend even to prosody: The contrastive reading of particular intonational patterns is strengthened when talkers use alternative forms for non-contrastive utterances but is weakened when talkers do not reliably use contrastive prosody to signal pragmatic alternatives (Kurumada, Brown, Bibyk, Pontillo, & Tanenhaus, 2014). Moreover, the absence of an expected beat gesture or pitch accent can increase the amplitude of the N400, an ERP component reflecting semantic processing difficulty (Wang & Chu, 2013). Together, these findings suggest that knowledge that the speaker sometimes uses a particular cue for prominence (e.g., beat gesture) may cause comprehenders to modify how they process language even when that cue is absent and even when the bottom-up input is otherwise the same. This, in turn, may inform the understanding of how comprehenders define the context of cue use within a discourse – that is, whether they sum cue use across talkers or whether they interpret cues differently depending on individual talkers' histories of cue use (a point we elaborate on in the *General discussion*).

Fig. 1 Predictions of how the absence of a beat gesture might be interpreted under data-explanation views when (**a**) the talker never produces beat gesture (corresponding to Experiment 2) and (**b**) the talker has previously produced beat gesture (corresponding to Experiment 1)

Present study

In the present study, we contrasted these competing hypotheses about how beat gesture and pitch accenting might be integrated and how this might affect discourse representation. We developed a novel paradigm that allowed us to present multimodal narratives while nevertheless exercising tight experimental control over both pitch accenting and beat gesture. We used cross-spliced audio recordings in which pitch accenting on critical words was manipulated while holding the rest of the auditory discourse constant (Fraundorf et al., 2010). Then, we synchronized this audio with different possible videos of a talker in which we manipulated the presence or absence of beat gesture. The talker's face was obscured, concealing any discrepancies in timing between lip movements and speech and allowing us to independently vary pitch accenting and beat gesture in a multimodal discourse in which the audio was otherwise identical across conditions. We then tested how beat gesture, contrastive pitch accenting, and their interaction affect subsequent recognition memory for the events of the discourse (paralleling the auditory-only paradigm introduced by Fraundorf et al., 2010). This paradigm allowed us to examine how beat gesture and contrastive pitch accenting influence memory for specific words within discourse, in contrast with previous work examining the effect of beat gesture on memory only in addition to contrastive pitch accenting (Kushch & Prieto, 2016; Llanes-Coromina et al., 2018) or only for individual words or entire sentences from discourse (Biau & Soto-Faraco, 2013; Biau et al., 2015; Feyereisen, 2006; Igualada et al., 2017; Kushch et al., 2018).

In Experiment 1, we independently manipulated the presence of contrastive pitch accenting and beat gesture on critical words. To preview, results supported a data-explanation view of language processing: When the speaker emphasized *some* critical words with beat gesture, the absence of beat gesture also became informative, such that another potential cue to discourse importance (contrastive pitch accents) no longer affected memory in cases where beat gesture was absent. Experiment 2 provided further support for this data-explanation interpretation by demonstrating that it was specifically the talker's use of beat gesture that drove this effect.

Experiment 1

Method

Participants

All participants in this and all subsequent studies were native English speakers aged 18 years or older with normal or corrected-to-normal hearing and vision.[2] We restricted our sample to native English speakers so that all participants were likely to have *a priori* knowledge of the use of pitch accenting and beat gesture in English (cf., Lee & Fraundorf, 2016).

[2] No information was collected concerning additional participant-level variables, such as participants' knowledge of other languages or musical ability. However, our within-subjects design controls for any potential individual differences due to these factors.

Participants were recruited via electronic and paper advertisements posted at the University of Pittsburgh and in the local community and were compensated with US$9 in cash for their participation. To ensure that our sample size was comparable to that of other similar studies and that an equal number of participants completed all 16 stimulus lists, we decided to recruit 32 individuals prior to running Experiment 1. Data from an additional four pilot participants and one participant for whom a technical error occurred were excluded from analyses.

Materials

Participants were presented with 32 prerecorded discourses (see Appendix A of Fraundorf, Benjamin, & Watson, 2010 for a complete list), which were presented as audiovisual clips of a talker telling a story. Each discourse consisted of a context passage followed by a continuation passage. The context passage established two pairs of contrasting alternatives; for instance, in example (7) below, *British* and *French* are one pair and *Malaysia* and *Indonesia* are the other. The continuation passage, such as (8) below, then mentioned one alternative from each pair (e.g., *British* from one pair and *Malaysia* from the other).

(7) Both the British and the French biologists had been searching Malaysia and Indonesia for the endangered monkeys.
(8) Finally, the (British/BRITISH) spotted one of the monkeys in (Malaysia/MALAYSIA) and planted a radio tag on it.

The discourses were constructed so that, across discourses, an equal number of continuation passages referred to the alternative that occurred first in the context passage – for example, *British* was mentioned before *French* in example (6) – compared to to the alternative that occurred second. Because there were two critical words per continuation passage, we fully counterbalanced whether the alternative mentioned in the continuation passage was first-mentioned or second-mentioned in the context passage for each of the two pairs in an orthogonal manner (i.e., across discourses, the patterns first-first, first-second, second-first, and second-second were all equally common). This counterbalancing prevented participants from anticipating which alternative from each pair would be mentioned in the continuation passage based on their ordering in the context passage.

The audio component of each discourse was taken from Fraundorf, Watson, and Benjamin (2010), Experiment 2 and was originally produced by a female research assistant who was a native speaker of the Inland North American English dialect (Labov, Ash, & Boberg, 2008). Each critical word in the continuation was produced with either a contrastive pitch accent (L+H*, capitalized in the example above) or presentational pitch accenting (H*, lowercase in the example above), counterbalanced across participants for each discourse. Acoustic analyses of the audio recordings (presented in Table 1 of Fraundorf et al., 2010) revealed that words receiving contrastive versus presentational accents differed significantly in intensity, duration, maximum pitch, difference between maximum and minimum pitch, and mean pitch on both the stressed syllable and on the entire critical word. To ensure that prosodic differences existed only on critical words, audio stimuli were created using cross-splicing to splice critical words into carrier passages that were constant across conditions. A structured debriefing taken from the original study queried whether participants noticed the splicing; no participants in any of the experiments reported here did.

To vary beat gesture in a paradigm that could be directly compared to the auditory-only presentation used by Fraundorf et al. (2010), we created new videos that were exactly matched to the existing audio clips. A female narrator was videotaped for all of the conditions (paralleling the use of a single speaker in the Fraundorf et al. audio materials). For each condition, the narrator first viewed the written text of each discourse and listened to the audio of that discourse in the corresponding pitch-accent condition. Then, the narrator was video recorded re-telling the discourse in tandem with the audio clip while producing the same pattern of pitch accenting. As the narrator spoke, she produced beat gestures in conjunction with critical words occurring in continuation passages of discourses or kept her hands still. Beat gestures consisted of a single downward flick of an open-palmed right hand, which is the most common (and prototypical) type of beat gesture produced spontaneously with speech (McNeill, 1992). The narrator produced beat gesture such that its stroke (downward motion) occurred in conjunction with the stressed syllable of critical words, [3] consistent with how beat gesture and speech prosody are aligned in natural conversation (Loehr, 2012).

The audio channel from the video was then removed and replaced with the original audio from the Fraundorf et al. (2010) materials, thereby excluding the narrator's speech. Beat gesture stroke onsets (i.e., apices) were temporally aligned with the onset of the stressed syllable of critical words in the original audio. This procedure allowed us to use audio materials from Fraundorf et al. (2010), which employed cross-splicing to hold constant the prosody of all but the critical words. Because a different audio track was used in the final

[3] To verify that this procedure was successful, after temporally aligning the stressed syllable of critical words in original audio clips with strokes of beat gesture in video clips in post-production (see below), we measured the difference between their onsets and offsets in a random selection of 25% of discourses. On average, the stressed syllable of critical words began 120 ms (*SD* = 90 ms) and ended 20 ms (*SD* = 120 ms) before beat gesture strokes, falling within the range in which temporal asynchronies between visual and auditory speech are not reliably perceived (Conrey & Pisoni, 2006).

Table 1 Summary of experimental design for Experiment 1

First critical word	Second critical word	Number per participant
Beat, contrastive accenting	Beat, contrastive accenting	4
Beat, contrastive accenting	No beat, presentational accenting	4
No beat, presentational accenting	Beat, contrastive accenting	4
No beat, presentational accenting	No beat, presentational accenting	4
No beat, contrastive accenting	No beat, contrastive accenting	4
Beat, presentational accenting	Beat, presentational accenting	4
No beat, contrastive accenting	Beat, presentational accenting	4
Beat, presentational accenting	No beat, contrastive accenting	4

materials than was recorded with video, it was impossible to perfectly synchronize lip movements with the final audio track. Thus, we edited the videos to pixelate and blur the narrator's face (see Fig. 2). This blurring also served to obscure any facial or head movements (which might have served as another cue to emphasis) and to direct participants' attention away from the face, in accordance with standard practice in gesture research. Participants were given a cover story stating that the speaker's face was blurred to obscure her identity.

On each of the two critical words, such as *British* and *Malaysia* in example (8) above, we independently manipulated beat gesture and pitch accenting, resulting in a 2 (beat gesture or no beat gesture) × 2 (contrastive or presentational accent) design for each word, with 16 critical words in each of these experimental cells.[4] (These trials constituted the entire experimental list; in this and Experiment 2, we did not include practice trials that might bias participants either to expect or not expect beat gestures in the remainder of the experiment.) Thus, the critical variables of beat gesture and pitch accenting were fully orthogonal within each critical word (see Fig. 2). It should be noted that there were minor contingencies across words, such that the manipulations on one critical word could not be fully independent of the manipulations on the other critical word in the same discourse. (That is, the manipulations on *British* were not fully independent of the manipulations on *Malaysia*.) Specifically, critical words with contrastive pitch accenting had slightly longer acoustic duration than those with presentational pitch accenting, and videos with beat gesture were slightly longer than videos without gesture.[5] This made it impossible in this paradigm to have contrastive accenting with beat gesture or presentational accenting without gesture on one critical word but not the other: Extending, on one critical word, the audio track but not the video track would misalign the speech and gesture for the other critical word. (By contrast, it was possible to have presentational accenting with beat gesture or contrastive accenting without gesture on both words because extending the audio track for one word and the video track for another word resulted in a clip of the same total length.) Thus, we counterbalanced pitch accenting and beat gesture *within* each critical word but not *across* both critical words, as outlined in Table 1. These minor contingencies are unlikely to affect participants' memory given prior evidence that contrastive pitch accenting on one critical word does not affect its impact on memory for another critical word, at least among typical young adults (Fraundorf et al., 2010, 2012). The assignment of discourses to pitch accenting and beat gesture conditions was counterbalanced using a Latin Square design, resulting in 16 presentation lists.

[4] Fraundorf et al. (2010) varied pitch accenting independently on two separate critical words to test whether the type of pitch accenting on one critical word affected processing of the other; however, no such effect was observed in their experiment or in the present study.

[5] Although it is not entirely clear why videos with beat gesture were slightly longer than videos without beat gesture, we believe, in light of the small magnitude of the duration difference, it is due to implicit production of greater prosodic prominence on critical words accompanied by beat gesture (see Krahmer & Swerts, 2007) rather than production of artificially slow beat gesture.

Procedure

The procedure of Experiment 1 paralleled that of Experiment 2 in Fraundorf, Watson, and Benjamin (2010). Participants were told that they would hear several discourses and that their memory for the discourses would be tested afterwards. Participants were told to try to remember as much as they could but were not given specific information about the format of the memory test. Participants performed the task on a computer running MATLAB with the Psychophysics Toolbox (Brainard, 1997; Kleiner et al., 2007) and the CogToolbox (Fraundorf et al., 2014).

Before beginning the task, participants heard a test sound that allowed them to adjust the volume to a comfortable level. Participants then began a study phase in which they heard all 32 discourses presented in randomized order with a 5-s delay between each one. After participants had listened to 16 of the discourses, they were informed that they were halfway through the task and were allowed to take a break if they wished.

After participants had listened to each discourse once, they entered the test phase. Complete discourses, including both the context and continuation passages, were presented on

Fig. 2 Schematic representation of the four possible combinations of beat gesture and pitch accenting on each critical word with face pixellation and blurring shown: (**a**) beat gesture with contrastive accenting; (**b**) no beat gesture with contrastive accenting; (**c**) beat gesture with presentational accenting; (**d**) no beat gesture with presentational accenting

screen one at a time as text with no accompanying sound or video. In continuation passages, critical words were replaced by underscores, as in example (9) below.

(9) Both the British and the French biologists had been searching Malaysia and Indonesia for the endangered monkeys. Finally, the ____[1]____ spotted one of the monkeys in ____[2]____ and planted a radio tag on it.

[1] (a) British (b) French [2] (a) Malaysia (b) Indonesia

Each trial in the test phase [1-2] probed memory for one contrast; thus, both contrasts in each discourse were tested. Memory for each contrast was probed by presenting the two alternatives corresponding to the highlighted contrast at the bottom of the screen. Participants pressed a key on the keyboard (a, b) to indicate their selection. Trials probing memory

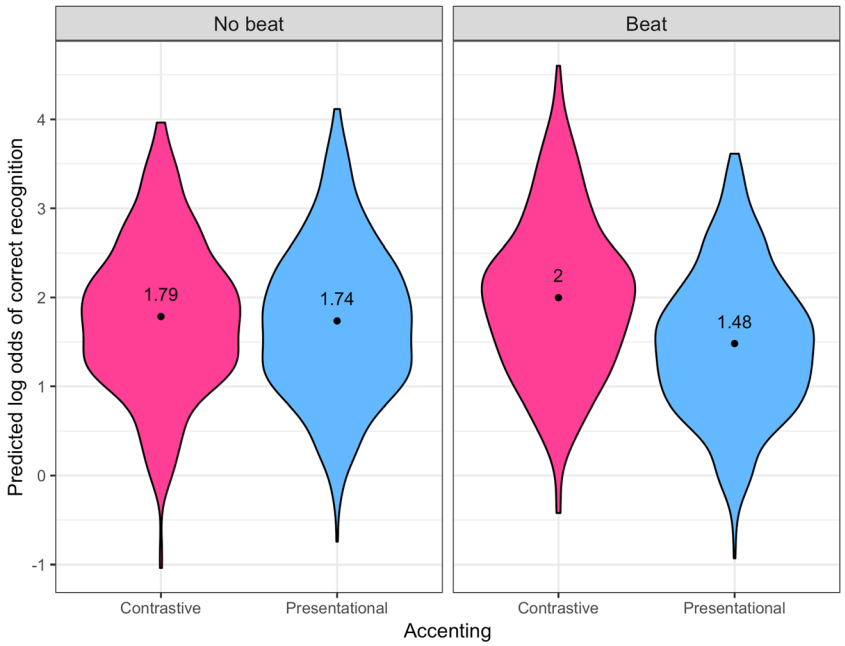

Fig. 3 Predicted log odds of recognition accuracy for critical words in Experiment 1 by beat gesture and pitch accenting (dots and values represent cell means)

for contrasts within the same discourse were separated by a 500-ms delay, and discourses were separated by a 1,000-ms delay.[6] The discourses were tested in a different random order than in the study phase.

Results

Recognition memory was analyzed as a function of pitch accenting and beat gesture in discourses. Accuracy for each cell of the experimental design is displayed in Fig. 3.

Data were analyzed using mixed effects logit models, which model the log odds of a correct response on each trial. Our model included fixed effects of pitch accenting, beat gesture, and their interaction, as well as crossed random intercepts for subjects and items. The addition of random slopes by subject or by item for any of the independent variables did not improve the fit of the model in likelihood-ratio tests (all *p*s > .19), so they were excluded. The model was fit in the R software package with Laplace estimation using the glmer() function of the *lme4* package (Bates, Mächler, Bolker, & Walker, 2015); intra-class correlation coefficients were estimated with package *sjstats* (Lüdecke, 2019). All fixed effects were coded as -0.5 and +0.5 to obtain estimates of the main effects analogous to those from an ANOVA.

Table 2 displays parameter estimates for the model.[7] Across gesture conditions, there was a main effect of contrastive pitch accenting on memory: The odds of correct recognition of critical words were approximately 1.33 times (95% CI: [1.04, 1.68]) greater when those words had contrastive pitch accenting than when they had presentational accenting in the spoken discourse, $p = .02$. By contrast, the presence of beat gesture in video discourses did not have a significant main effect on subsequent recognition accuracy, $p = .87$.

However, this effect of contrastive accenting was largely qualified by beat gesture; the effect of contrastive accenting on the odds of correct subsequent recognition was 1.61 times (95% CI: [1.00; 2.60]) greater when beat gesture was present than absent in a video discourse, $p = .05$ (see Fig. 2). Simple main effect analyses by beat gesture revealed that when critical words were accompanied by beat gesture in a video discourse, the odds of correct subsequent recognition were 1.67 times (95% CI: [1.19; 2.35]) greater when these words had contrastive accents than when they had presentational accents, $p = .003$. By contrast, when critical words were unaccompanied by beat gesture in a video discourse, the odds of correct subsequent recognition were only 1.04 times (95% CI: [0.75; 1.46]) greater with contrastive accenting than presentational accenting, $p = .80$.

Table 2 Fixed effect estimates (top) and variance estimates (bottom) in log odds for multi-level logit model of correct recognition in Experiment 1 (observations = 2,040)

Fixed effect	Coefficient	*SE*	Wald *z*	*p*
Intercept	1.82	0.18	10.29	< .001
Beat gesture	0.02	0.12	0.17	.87
Contrastive accenting	-0.28	0.12	-2.31	.02
Beat gesture x contrastive accenting	0.48	0.24	1.95	.05
Random effect	s^2	ICC		
Participant	0.41	.14		
Item	0.62	.09		

Simple main effect analyses by pitch accent did not reveal a significant effect of beat gesture when critical words were accompanied by contrastive accenting, $p = .27$, or presentational accenting, $p = .14$. Taken together, these results suggest that beat gesture does not directly affect memory for discourse, with or without contrastive accenting; rather, it modulates the effectiveness of pitch accenting.

Discussion

In Experiment 1, in which beat gesture accompanied some critical words but not others, its absence negated the effect of pitch accenting on memory for discourse. Specifically, when critical words were accompanied by beat gesture, critical words were better remembered when pitch accenting was contrastive rather than presentational. However, when beat gesture was absent, we found no significant effect of pitch accent type.

This pattern is noteworthy because the auditory input of our no-gesture trials was identical to previous studies that found an effect of contrastive accenting when no gestures were ever used (Fraundorf et al., 2010, 2012). This result suggests that the absence of a cue that sometimes occurs (i.e., beat gesture) can negate the effects of another cue (in this case, pitch accenting), inconsistent with a purely bottom-up view of discourse processing. Rather, this finding is consistent with data-explanation views, in which the absence of a cue may suggest unimportance if they believe that the talker *would* have produced the cue *had* the information been important. In the case of Experiment 1, for instance, if comprehenders interpret the talker's beat gesture as indicating that a critical word is important, they are likely to pay less attention to critical words unaccompanied by beat gesture and to

[6] Data for one discourse (trial 5, contrastive accenting on first critical word, beat gesture on second critical word) was excluded due to a technical issue with the video discovered after data collection.

[7] Because our hypothesis is that comprehenders adapt to the talker's use of beat gesture, another test could be whether these effects become stronger over the course of the experiment. However, we did not find a significant interaction of trial number with any other variable in this experiment (all *p*s ≥ .22) or the subsequent one ($p = .30$). Nevertheless, we argue that such an effect may have been unlikely to emerge. This trial-number analysis tests for a linear trend, in which the effect of gesture (or pitch accent) changes steadily from one discourse to the next. However, the critical difference between Experiments 1 and 2 was whether the talker uses beat gesture at all, which can be learned from even a single exposure to beat gesture (i.e., one-trial learning) and would therefore not necessarily result in a gradual, linear change over time.

consequently disregard pitch accenting on those words, eliminating any mnemonic difference between contrastive and presentational pitch accenting when beat gesture is absent. Thus, this finding suggests that comprehenders take talkers' beat gesture production into account by creating a mental model of their communicative intent that influences their attention to and memory for discourse.

In this account, the result observed in no-gesture trials arises specifically because the talker uses beat gesture on some *other* trials, not because pitch accenting effects only *ever* obtain in the presence of beat gesture. Evidence for this claim comes from previous experiments by Fraundorf et al. (2010, 2012) that used exclusively auditory presentation. In those studies, contrastive and presentational pitch accenting differed in their effects on discourse memory even though no gestures ever occurred. However, an alternate explanation of the difference between the present results and those of Fraundorf et al. (2010, 2012) is simply that the present experiments use video materials, rather than the presence of beat gesture per se. To address this question, we conducted a second experiment in which we presented video discourses in which beat gesture was *never* present. We hypothesized that when the talker never uses beat gesture, critical words with contrastive pitch accenting would be recognized more accurately than critical words with presentational pitch accenting even in the absence of gesture. Thus, we anticipated that the results of Experiment 2 would provide additional support for the data-explanation account by demonstrating that comprehenders adapt their expectations to the types of cues produced in the current context.

Experiment 2

Method

Participants

Prior to running Experiment 2, we decided to recruit 32 new participants meeting the criteria outlined for Experiment 1 in order to have a sample size comparable to that of Experiment 1. Data from an additional four pilot participants and two participants for whom technical errors occurred were excluded from analyses.

Materials

Participants were presented with the same 32 prerecorded discourses from Experiment 1. We also used the same videos in Experiment 2, except that we included only those videos that did not contain beat gesture. Because the goal of Experiment 2 was to replicate the no-gesture conditions of Experiment 1 except in a context where no beat gestures were ever present, the manipulation of pitch accenting proceeded exactly in Experiment 1: Both critical words had either contrastive or presentational accenting (see Table 3). Assignment of discourses to conditions was counterbalanced across participants using a Latin Square design, resulting in two presentation lists and 32 critical words in each of the two experiental conditions.

Procedure

The procedure of Experiment 2 was identical to that of Experiment 1.

Results

Recognition memory was analyzed as a function of pitch accenting and beat gesture in discourses. Accuracy is displayed in Fig. 4.

The mixed-effect logit model of recognition accuracy included a fixed effect of pitch accenting as well as crossed random intercepts for subjects and items. The addition of random slopes for pitch accenting by item improved the fit of the model in likelihood-ratio tests ($p > .001$), so they were included in the final model. As in Experiment 1, the fixed effect was coded as -0.5 and +0.5.

Table 4 displays parameter estimates for the model. In this experiment, in which beat gesture was never present, contrastive pitch accenting yielded better memory for discourse than presentational pitch accenting, even in the absence of gesture: The odds of subsequent recognition of critical words were approximately 1.33 times (95% CI: [1.03, 1.72]) greater when those words had been heard in the spoken discourse with contrastive accenting than with presentational accenting, $p = .03$. These results contrast with those of Experiment 1, in which trials without beat gesture showed no difference in recognition accuracy as a function of pitch accenting.

Discussion

Experiment 2 showed that, when beat gesture is never present, critical words from a discourse that were originally heard with contrastive pitch accenting are later remembered more accurately than words originally heard with presentational pitch accenting. Critically, this was the case even though the bottom-up input was exactly the same as in trials without beat gesture in Experiment 1. Thus, this finding suggests that the interactive effect of beat gesture and pitch accenting on memory in Experiment 1 was driven by comprehenders' adaptation to the presence of beat gesture on some trials rather than the use of video in and of itself.Given that the only difference between these two experiments was whether beat gesture was present in some cases (Experiment 1) or completely absent (Experiment 2), these diverging results provide additional evidence that available cues to prominence shape

Table 3 Summary of experimental design for Experiment 2

First critical word	Second critical word	Number per participant
No beat, contrastive accenting	No beat, contrastive accenting	16
No beat, presentational accenting	No beat, presentational accenting	16

comprehenders' inferences about talkers' intentions, in turn influencing their discourse interpretation. More specifically, when beat gesture is never produced in a particular context, comprehenders may interpret information with contrastive pitch accenting as important to the talker and therefore may devote greater attentional and memory resources to it. When talkers sometimes produce beat gesture, however, comprehenders appear to consider pitch accenting only when the talker has also signaled prominence with a beat gesture. Together, these results are consistent with data-explanation views of language processing, demonstrating that cue integration is driven by sensitivity to talkers' communicative intent and, consequently, consideration of the cues that talkers use in a particular context.

Another contribution of Experiment 2 stems from the fact that the memory effects of pitch accenting in Experiment 2 are analogous to those shown in previous experiments (Experiments 1 and 2 of Fraundorf et al., 2010, 2012) that used a similar paradigm but in which discourses were presented only as audio. Thus, Experiment 2 demonstrates that the effect of pitch accenting on discourse memory can generalize to more naturalistic audiovisual stimuli, which are more similar to most person-to-person conversational contexts than purely auditory stimuli (see also Kushch & Prieto, 2016; Llanes-Coromina et al., 2018).

General discussion

We examined how beat gesture and pitch accenting are integrated in discouse comprehension and how they affect subsequent memory for a discourse. In Experiment 1, the talker sometimes emphasized critical words with a beat gesture. Here, contrastive (L+H*) pitch accenting yielded memory superior to presentational (H*) pitch accenting only for those words emphasized by a beat gesture, consistent with (and expanding upon) the findings of other work using a non-factorial design (Kushch & Prieto, 2016; Llanes-Coromina et al., 2018). By contrast, in Experiment 2, the talker never produced beat gesture; here, contrastive accenting enhanced memory even in the absence of beat gesture, consistent with experiments using an auditory-only paradigm in which gesture was necessarily absent (Fraundorf et al., 2010, 2012).

Together, the results of Experiments 1 and 2 imply that even the *absence* of one cue (beat gesture) can become meaningful if there is reason to believe the talker would have produced the cue had the information been important. This favors data-explanation views of language processing in which the goal of comprehension is to model the talker's communicative intent.

More broadly, the current work, along with Krahmer and Swerts (2007; Kushch & Prieto, 2016; Llanes-Coromina et al.,

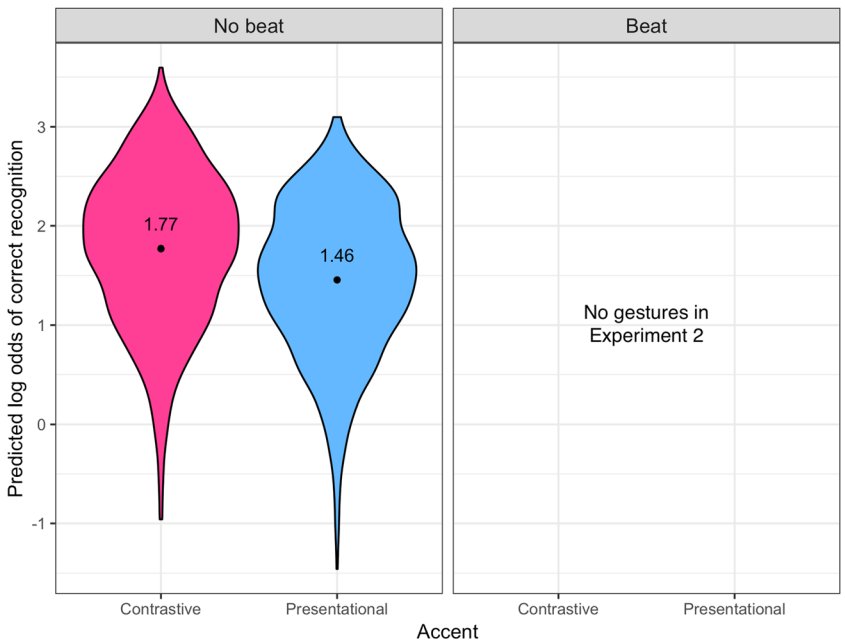

Fig. 4 Predicted log odds of recognition accuracy for critical words in Experiment 2 by pitch accenting (dots and values represent cell means)

Table 4 Fixed effect estimates (top) and variance estimates (bottom) in log odds for multi-level logit model of correct recognition in Experiment 2 (observations = 2112)

Fixed effect	Coefficient	*SE*	Wald *z*	*p*
Intercept	1.68	0.16	10.83	< .001
Contrastive accenting	-0.29	0.13	-2.21	.03
Random effect	s^2	ICC		
Participant	0.39	.09		
Item	0.49	.12		
Item x contrastive accenting	0.25	—		

2018) indicates that comprehenders integrate beat gesture and pitch accenting, affecting their interpretation of – and memory for – discourses containing these cues. Furthermore, these results suggest that interpretation of beat gesture and pitch accenting is influenced by the co-occurence pattern of these cues within a particular linguistic context. Cue co-occurrence patterns are important because discourse often contains several different types of cues produced in various modalities that sometimes overlap in timing and/or function (e.g., So, Kita, & Goldin-Meadow, 2009). Thus, to understand how language is interpreted in a naturalistic context, it is crucial to understand the impact of cue co-occurrence across modalities and in more elaborate discourses. However, most research examining the influence of pitch accenting on memory for discourse has used purely auditory stimuli (but see Kushch & Prieto, 2016; Llanes-Coromina et al., 2018). Moreover, most past work establishing that beat gesture can enhance later memory (Levantinou & Navarretta, 2016; Morett, 2014; So et al., 2012) has used lists of unrelated words or thematically-related sentences rather than a semantically richer discourse. Here, we examined the effects of both beat gesture and pitch accenting on memory for contrasting alternatives within a complex discourse. Consistent with Kushch and Prieto (2016) and Llanes-Coromina et al. (2018), who used a similar paradigm, we found that beat gesture and contrastive pitch accenting affect memory even for complex, multimodal discourses.

Inferences about communicative intent

More generally, the present results support emerging data-explanation views of language processing (or of cognition more broadly; Clark, 2013), in which the goal of language comprehension is to infer talkers' underlying intentions (Farmer, Brown, & Tanenhaus, 2013). In these views, processing of linguistic input received is influenced by the input that could have been received had the talker's possible communicative intentions differed (Rohde & Kurumada, 2018). Here, for instance, we observe that memory for spoken discourse is guided not just by the cues present in the signal at any given moment – such as beat gesture or contrastive pitch accenting – but by other cues that the talker could have produced had the underlying communicative intent been different. The findings of the present work are consistent with research demonstrating that availability of alternatives with contrasting meanings affects discourse interpretation in general (Bergen, Goodman, & Levy, 2012; Degen & Tanenhaus, 2016), as well as interpretation of contrastive pitch accenting in particular (Kurumada et al., 2014). The present research builds on these findings by suggesting that a previous pattern of productions indicating that the talker *would* have produced a beat gesture *had* a critical fact been important affects comprehenders' interpretation of any co-occuring pitch accenting.

While it is possible that such comparison against other possible linguistic input may sometimes be a conscious strategy, it need not be a conscious or even effortful process. Other work has identified how rational inferences about the source of perceptual input could be approximated in ways that are neurally and computationally plausible (for further discussion, see Clark, 2013). Thus, in many cases, these comparisons may be an automatic, implicit process of the language comprehension system. Although the current study does not directly address whether this comparison is explicit or implicit, future research should address this important question, providing further insight into how the presence – and absence – of cues to prominence affects spoken discourse processing.

Contextual and structural influences on cue integration

While the present work showed that comprehenders are sensitive to the set of cues produced in the current discursive context, a further question – beyond the scope of the current project – is is exactly *how* the context of cue use in discourse is specified. For instance, one possibility is that this context encompasses the communicative preferences of specific *talkers*, such that discourse produced by a talker who sometimes produces beat gesture is interpreted and remembered differently than discourse produced by a talker who never (or always) produces beat gesture (Chu et al., 2014). On the other hand, because talkers from a similar background are likely to share linguistic preferences (e.g., Eckert & Rickford, 2001), discursive context might be specified at a more general level, such as the entire experiment, so that patterns of cues produced by individual talkers within the experiment would not differentially affect discourse interpretation and memory. Future research could attempt to distinguish between these possibilities by examining how beat gesture and pitch accenting affect memory in multi-talker paradigms in which different talkers within an experiment produce different patterns of beat gesture and pitch accenting, as has been done for speech perception (e.g., Kraljic & Samuel, 2005, 2006; Munson, 2011;

Nygaard & Pisoni, 1998) and syntactic variability (Kamide, 2012).

Bottom-up influences on cue integration

The findings of this research are inconsistent with a purely bottom-up account in which the mere presence of either beat gesture or contrastive pitch accenting intrinsically functions as a "processing instruction" that attracts attention to the speech stream, enhancing memory for any words accompanied by these cues. According to such accounts, an effect of contrastive pitch accenting should have been present for discourses in the no-gesture condition in Experiment 1. Considered in conjunction with the effect of contrastive pitch accenting observed when beat gesture was never present in Experiment 2, the absence of an effect of contrastive pitch accenting in the no-gesture condition in Experiment 1 indicates that information about the circumstances under which these cues are produced is taken into account during discourse interpretation. Thus, although the processing instructions account may provide a plausible explanation for how some individual linguistic cues, such as grammatical focus, are interpreted, it does not appear to explain how multiple linguistic cues are integrated within discourse, limiting its scope and applicability.

More generally, perceptual salience alone cannot fully explain the observed effects. The perceptual input on no-gesture trials was the same between Experiment 1 and Experiment 2, yet only Experiment 2 showed an effect of contrastive pitch accenting on no-gesture trials. Several prior findings also suggest that perceptual salience alone cannot account for the effects of linguistic cues on memory: Highlighting information through grammatical focus and discursive context benefits memory even though perceptual characteristics are not affected (Sturt, Sanford, Stewart, & Dawydiak, 2004), increasing the perceptual salience of text through capitalization does not necessarily enhance semantic processing (Kamas, Reder, & Ayers, 1996), and increasing the perceptual salience of speech via pitch accenting increases rejection of certain alternatives but not others (Fraundorf, Watson, & Benjamin, 2010, Experiment 3). It is important to note that although perceptual characteristics alone are insufficient to explain the results of current and previous research, they are nonetheless necessary to explain these results, including the finding from the current study that pitch accenting is subordinate to beat gesture when talkers produce both cues. Thus, it is probably no coincidence that more visually or acoustically prominent devices are used to highlight important information. Indeed, one interesting aspect of the present study is that beat gesture conditioned the effect of pitch accenting rather than the reverse, suggesting that beat gesture may serve as a stronger cue to prominence than pitch accenting. This may be the case because visual cues to prominence, such as beat gesture, are more salient and/or less common than auditory cues to prominence, such as pitch accenting.

Similarly, the present findings cannot be reduced to the *blocking effect* (Kamin, 1969) observed in classical conditioning, in which the association between a previously-trained conditioned stimulus and an unconditioned stimulus blocks a second association from being learned. For example, mice that already salivate upon hearing a tone associated with food will not later learn to also associate food with a flash of light presented at the same time as the tone. Applied to linguistic cue integration, we might see a blocking effect if the presence of one cue (e.g., beat gesture) blocked the effect of a second cue (e.g., contrastive pitch accenting) on memory. However, we observed the opposite pattern; it was not the the *presence* of beat gesture that negated contrastive pitch accenting's effect on memory, but rather its *absence*. Furthermore, the classical conditioning experiments in which the blocking effect was demonstrated involved newly learned associations, whereas the current study did not provide any within-experiment training on the meaning of beat gesture or contrastive prosody and relied on participants' a priori knowledge about these cues.

Implications and extensions

The findings of the current research have several implications for how beat gesture and contrastive pitch accenting convey information. Consistent with the findings of previous research (Krahmer & Swerts, 2007; Kushch & Prieto, 2016; Llanes-Coromina et al., 2018), this work demonstrates that both beat gesture and contrastive pitch accenting can enhance listeners' memory for a discourse. However, they do not address specifically what aspects of the discourse representation it enhances. Nevertheless, previous research suggests that contrastive pitch accenting enhances memory by ruling out the specific contrastive alternatives mentioned in the discourse. Namely, in a variant of the paradigm employed in the present study (Fraundorf et al. 2010, Experiment 3; see also Spalek et al., 2014), contrastive pitch accenting specifically facilitated the rejection of alternatives from the contrast set established in the discourse (e.g., the *French* scientists in example 3) rather than wholly unmentioned items (e.g., *German* scientists). Future research could examine whether beat gesture functions in a similar manner; doing so would probe the degree of similarity between the cognitive functions of beat gesture and contrastive pitch accenting.

Because our primary question of interest in this work was how the absence of beat gesture affects interpretation of pitch accenting, we did not investigate how pitch accenting is interpreted when beat gesture is always present. Although the results of such a study would not directly address whether cue integration proceeds in a top-down or bottom-up manner, which was our primary theoretical focus in this work, they would reveal whether the effect of contrastive pitch accenting

observed when beat gesture was present in Experiment 1 generalizes to contexts in which beat gesture is always present or is specific to cases in which beat gesture is present only some of the time. In a similar vein, the results of a study examining how manipulating beat gesture affects memory for discourse when pitch accenting is held constant (e.g., when all critical words have presentational pitch accenting) would provide further evidence that beat gesture enhances memory for information in discourse, even though it would not directly address how multiple cues are integrated, another primary focus of this work.

Conclusion

The current research demonstrates that the presence – and absence – of beat gesture alters the effect of contrastive pitch accenting on memory for discourse. The results revealed that contrastive pitch accenting enhances memory for critical information in discourse when it occurs when beat gesture is never present or in conjunction with beat gesture within contexts in which beat gesture is sometimes present. However, this effect disappears when contrastive pitch accenting occurs in the absence of beat gesture in contexts in which beat gesture is sometimes present. These findings support data-explanation views of language processing in which the cues that a speaker *could* produce influence comprehension even in the absence of those cues.

Open Practices Statement All data and analysis scripts are publically accessible via the Open Science Framework at the following URL: https://osf.io/kxsrf/. The experiments reported in this paper were not preregistered.

Author Note We thank Alba Tuninetti for assistance with video recording, Laurel Brehm for comments on an earlier version of this manuscript, and Emalee Dauginikas, Rachel Peters, Anisah Rafi, and Carmen Sepulveda for assistance with data collection.

References

Allen, J. S., Miller, J. L., & DeSteno, D. (2003). Individual talker differences in voice-onset-time. *Journal of the Acoustical Society of America*, *113*, 544–552. https://doi.org/10.1121/1.1528172

Almor, A., & Eimas, P. D. (2008). Focus and noun phrase anaphors in spoken language comprehension. *Language and Cognitive Processes*, *23*, 201–225. https://doi.org/10.1080/01690960701330936

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48. https://doi.org/10.18637/jss.v067.i01

Bergen, L., Goodman, N., & Levy, R. (2012). That's what she (could have) said: How alternative utterances affect language use. In N. Miyake, D. Peebles, & R. Cooper (Eds.), Proceedings of CogSci 2012 (pp. 120–125). Austin, TX: Cognitive Science Society.

Biau, E., & Soto-Faraco, S. (2013). Beat gestures modulate auditory integration in speech perception. *Brain and Language*, *124*, 143–152. https://doi.org/10.1016/j.bandl.2012.10.008

Biau, E., Torralba, M., Fuentemilla, L., de Diego Balaguer, R., & Soto-Faraco, S. (2015). Speaker's hand gestures modulate speech perception through phase resetting of ongoing neural oscillations. *Cortex*, *68*, 76–85. https://doi.org/10.1016/j.cortex.2014.11.018

Birch, S. L., Albrecht, J. E., & Myers, J. L. (2000). Syntactic focusing structures influence discourse processing. *Discourse Processes*, *30*, 285–304. https://doi.org/10.1207/S15326950dp3003_4

Birch, S. L., & Garnsey, S. M. (1995). The effect of focus on memory for words in sentences. *Journal of Memory and Language*, *34*, 232–267. https://doi.org/10.1006/jmla.1995.1011

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*, 433–436. https://doi.org/10.1163/156856897X00357

Braun, B., & Tagliapietra, L. (2010). The role of contrastive intonation contours in the retrieval of contextual alternatives. *Language and Cognitive Processes*, *25*, 1024–1043. https://doi.org/10.1080/01690960903036836

Brown, G. (1983). Prosodic structure and the given/new distinction. In A. Cutler & D. R. Ladd (Eds.), Prosody: models and measurements (pp. 67–77). Berlin: Springer.

Chu, M., Meyer, A., Foulkes, L., & Kita, S. (2014). Individual differences in frequency and saliency of speech-accompanying gestures: The role of cognitive abilities and empathy. *Journal of Experimental Psychology: General*, *143*, 694. https://doi.org/10.1037/a0033861

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*, 181–204. https://doi.org/10.1017/S0140525X12000477

Cohen, R. L., & Otterbein, N. (1992). The mnemonic effect of speech gestures: pantomimic and non-pantomimic gestures compared. *European Journal of Cognitive Psychology*, *4*, 113. https://doi.org/10.1080/09541449208406246

Conrey, B., & Pisoni, D. B. (2006). Auditory-visual speech perception and synchrony detection for speech and nonspeech signals. *The Journal of the Acoustical Society of America*, *119*, 4065–4073. https://doi.org/10.1121/1.2195091

Cook, S. W., & Tanenhaus, M. K. (2009). Embodied communication: speakers' gestures affect listeners' actions. *Cognition*, *113*, 98–104. https://doi.org/10.1016/j.cognition.2009.06.006

Degen, J., & Tanenhaus, M. K. (2016). Availability of alternatives and the processing of scalar implicatures: A visual world eye-tracking study. *Cognitive Science*, *40*, 172–201. https://doi.org/10.1111/cogs.12227

Eckert, P., & Rickford, J. R. (2001). Style and sociolinguistic variation. Cambridge, UK: Cambridge University Press.

Esteve-Gibert, N., & Prieto, P. (2013). Prosodic structure shapes the temporal realization of intonation and manual gesture movements. *Journal of Speech, Language, and Hearing Research*, *56*, 850–864. https://doi.org/10.1044/1092-4388(2012/12-0049)

Esteve-Gibert, N., Prieto, P., & Pons, F. (2015). Nine-month-old infants are sensitive to the temporal alignment of prosodic and gesture prominences. *Infant Behavior and Development*, *38*, 126–129. https://doi.org/10.1016/j.infbeh.2014.12.016

Farmer, T. A., Brown, M., & Tanenhaus, M. K. (2013). Prediction, explanation, and the role of generative models in language processing. *Behavioral and Brain Sciences*, *36*, 211–212. https://doi.org/10.1017/S0140525X12002312

Feyereisen, P. (2006). Further investigation on the mnemonic effect of gestures: Their meaning matters. *European Journal of Cognitive Psychology*, *18*, 185–205. https://doi.org/10.1080/09541440540000158

Fraundorf, S. H., Diaz, M. I., Finley, J. R., Lewis, M. L., Tooley, K. M., Isaacs, A. M., … Brehm, L. (2014). *CogToolbox for MATLAB*. Retrieved from http://www.scottfraundorf.com/cogtoolbox.html

Fraundorf, S. H., Watson, D. G., & Benjamin, A. S. (2010). Recognition memory reveals just how CONTRASTIVE contrastive accenting really is. *Journal of Memory and Language*, *63*, 367–386. https://doi.org/10.1016/j.jml.2010.06.004

Fraundorf, S. H., Watson, D. G., & Benjamin, A. S. (2012). The effects of age on the strategic use of pitch accents in memory for discourse: A processing-resource account. *Psychology and Aging*, *27*, 88–98. https://doi.org/10.1037/a0024138

Givón, T. (1992). The grammar of referential coherence as mental processing instructions. *Linguistics*, *30*, 5–56. https://doi.org/10.1515/ling.1992.30.1.5

Gotzner, N., & Spalek, K. (2019). The life and times of focus alternatives: Tracing the activation of alternatives to a focused constituent in language comprehension. *Language and Linguistics Compass*, *13*(2), e12310. https://doi.org/10.1111/lnc3.12310

Hirata, Y., Kelly, S. D., Huang, J., & Manansala, M. (2014). Effects of hand gestures on auditory learning of second-language vowel length contrasts. *Journal of Speech, Language, and Hearing Research*, *57*, 2090–2101. https://doi.org/10.1044/2014_JSLHR-S-14-0049

Hostetter, A. B., & Alibali, M. W. (2010). Language, gesture, action! A test of the Gesture as Simulated Action framework. *Journal of Memory and Language*, *63*, 245–257. https://doi.org/10.1016/j.jml.2010.04.003

Husband, E. M., & Ferreira, F. (2016). The role of selection in the comprehension of focus alternatives. *Language, Cognition and Neuroscience*, *31*, 217–235. https://doi.org/10.1080/23273798.2015.1083113

Igualada, A., Esteve-Gibert, N., & Prieto, P. (2017). Beat gestures improve word recall in 3-to 5-year-old children. *Journal of Experimental Child Psychology*, *156*, 99–112. https://doi.org/10.1016/j.jecp.2016.11.017

Ito, K., Jincho, N., Minai, U., Yamane, N., & Mazuka, R. (2012). Intonation facilitates contrast resolution: Evidence from Japanese adults and 6-year olds. *Journal of Memory and Language*, *66*, 265–284. https://doi.org/10.1016/j.jml.2011.09.002

Ito, K., & Speer, S. R. (2008). Anticipatory effects of intonation: Eye movements during instructed visual search. *Journal of Memory and Language*, *58*, 541–573. https://doi.org/10.1016/j.jml.2007.06.013

Kaland, C., Krahmer, E., & Swerts, M. (2011). Contrastive intonation: Speaker-or listener-driven? *Proceedings of the 17th Annual Conference International Conference on Phonetic Sciences*, 1006–1009. Hong Kong: City University of Hong Kong.

Kamas, E. N., Reder, L. M., & Ayers, M. S. (1996). Partial matching in the Moses illusion: Response bias not sensitivity. *Memory & Cognition*, *24*, 687–699. https://doi.org/10.3758/BF03201094

Kamide, Y. (2012). Learning individual talkers' structural preferences. *Cognition*, *124*, 66–71. https://doi.org/10.1016/j.cognition.2012.03.001

Kamin, L. J. (1969). Predictability, surprise, attention, and conditioning. In B. A. Campbell & R. M. Church (Eds.), Punishment and aversive behavior (pp. 279–296). New York: Appleton-Century-Crofts.

Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., & Broussard, C. (2007). *What's new in Psychtoolbox-3*. Presented at the 30th European Conference on Visual Perception (ECVP 2007), Arezzo, Italy.

Krahmer, E., & Swerts, M. (2007). The effects of visual beats on prosodic prominence: Acoustic analyses, auditory perception and visual perception. *Journal of Memory and Language*, *57*, 396–414. https://doi.org/10.1016/j.jml.2007.06.005

Kraljic, T., & Samuel, A. G. (2005). Perceptual learning for speech: Is there a return to normal? *Cognitive Psychology*, *51*, 141–178. https://doi.org/10.1016/j.cogpsych.2005.05.001

Kraljic, T., & Samuel, A. G. (2006). Generalization in perceptual learning for speech. *Psychonomic Bulletin & Review*, *13*, 262–268. https://doi.org/10.3758/BF03193841

Kurumada, C., Brown, M., Bibyk, S., Pontillo, D. F., & Tanenhaus, M. K. (2014). Is it or isn't it: Listeners make rapid use of prosody to infer speaker meanings. *Cognition*, *133*, 335–342. https://doi.org/10.1016/j.cognition.2014.05.017

Kushch, O., Igualada, A., & Prieto, P. (2018). Prominence in speech and gesture favour second language novel word learning. *Language, Cognition and Neuroscience*, 992–1004. https://doi.org/10.1080/23273798.2018.1435894

Kushch, O., & Prieto, P. (2016). The effects of pitch accentuation and beat gestures on information recall in contrastive discourse. In J. Barnes, A. Brugos, S. Shattuck-Hufnagel, & N. Veilleux (Eds.), Speech prosody 2016 (pp. 922–925). Baixas, France: International Speech Communication Association.

Labov, W., Ash, S., & Boberg, C. (2008). The atlas of North American English: Phonetics, phonology and sound change. Berlin: Walter de Gruyter.

Ladd, D. R. (2008). Intonational phonology. New York: Cambridge University Press.

Lee, E.-K., & Fraundorf, S. H. (2016). Effects of contrastive accents in memory for L2 discourse. *Bilingualism: Language and Cognition*, *20*, 1–17. https://doi.org/10.1017/S1366728916000638

Lee, E.-K., & Snedeker, J. (2016). Effects of contrastive accents on children's discourse comprehension. *Psychonomic Bulletin & Review*, *23*, 1589–1595. https://doi.org/10.3758/s13423-016-1069-7

Leonard, T., & Cummins, F. (2011). The temporal relation between beat gestures and speech. *Language and Cognitive Processes*, *26*, 1457–1471. https://doi.org/10.1080/01690965.2010.500218

Levantinou, E. I., & Navarretta, C. (2016). An investigation of the effect of beat and iconic gestures on memory recall in L2 speakers. In *Proceedings from the 3rd European Symposium on Multimodal Communication* (pp. 32–37). Linköping, Sweden: Linköping University Electronic Press.

Llanes-Coromina, J., Vilà-Giménez, I., Kushch, O., Borràs-Comes, J., & Prieto, P. (2018). Beat gestures help preschoolers recall and comprehend discourse information. *Journal of Experimental Child Psychology*, *172*, 168–188. https://doi.org/10.1016/j.jecp.2018.02.004

Loehr, D. P. (2012). Temporal, structural, and pragmatic synchrony between intonation and gesture. *Laboratory Phonology*, *3*, 71–89. https://doi.org/10.1515/lp-2012-0006

Lüdecke, D. (2019). *Statistical functions for regression models [R package version 0.17.3]*. Retrieved from https://CRAN.R-project.org/package=sjstats

McNeill, D. (1992). Hand and mind. Chicago: University of Chicago Press.

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

McNeill, D. (2006). Gesture: A psycholinguistic approach. *The Encyclopedia of Language and Linguistics*, 58–66.

Morett, L. M. (2014). When hands speak louder than words: The role of gesture in the communication, encoding, and recall of words in a novel second language. *The Modern Language Journal*, *98*, 834–853. https://doi.org/10.1111/modl.12125

Munson, C. M. (2011). *Perceptual learning in speech reveals pathways of processing. Unpublished doctoral dissertation*. University of Iowa, Iowa City, IA.

Nygaard, L. C., & Pisoni, D. B. (1998). Talker-specific learning in speech perception. *Perception & Psychophysics*, *60*, 355–376. https://doi.org/10.3758/BF03206860

Pierrehumbert, J., & Hirschberg, J. (1990). The meaning of intonational contours in the interpretation of discourse. In P. R. Cohen (Ed.), *Intentions in communication* (pp. 271–311). Cambridge, MA: MIT Press.

Repp, B. H. (1982). Phonetic trading relations and context effects: New experimental evidence for a speech mode of perception.

Psychological Bulletin, *92*, 81. https://doi.org/10.1037/0033-2909.92.1.81

Rohde, H., & Kurumada, C. (2018). Alternatives and inferences in the communication of meaning. In K. D. Federmeier & D. G. Watson (Eds.), The psychology of learning and motivation: Current topics in language (Vol. 68, pp. 215–262). San Diego: Academic Press.

Roustan, B., & Dohen, M. (2010a). Co-production of contrastive prosodic focus and manual gestures: Temporal coordination and effects on the acoustic and articulatory correlates of focus. In M. Hasegawa-Johnson (Ed.), Speech prosody 2010. Baixas, France: International Speech Communication Association.

Roustan, B., & Dohen, M. (2010b). Gesture and speech coordination: The influence of the relationship between manual gesture and speech. In M. Hasegawa-Johnson (Ed.), Speech prosody 2010. Baixas, France: International Speech Communication Association.

Rusiewicz, H. L., Shaiman, S., Iverson, J. M., & Szuminsky, N. (2013). Effects of prosody and position on the timing of deictic gestures. *Journal of Speech, Language, and Hearing Research*, *56*, 458–470. https://doi.org/10.1044/1092-4388(2012/11-0283)

Rusiewicz, H. L., Shaiman, S., Iverson, J. M., & Szuminsky, N. (2014). Effects of perturbation and prosody on the coordination of speech and gesture. *Speech Communication*, *57*, 283–300. https://doi.org/10.1016/j.specom.2013.06.004

Sanford, A. J., Sanford, A. J., Molle, J., & Emmott, C. (2006). Shallow processing and attention capture in written and spoken discourse. *Discourse Processes*, *42*, 109–130. https://doi.org/10.1207/s15326950dp4202_2

Schwarzschild, R. (1999). GIVENness, AvoidF and other constraints on the placement of accent. *Natural Language Semantics*, *7*, 141–177. https://doi.org/10.1023/A:1008370902407

Shattuck-Hufnagel, S., & Ren, A. (2018). The prosodic characteristics of non-referential co-speech gestures in a sample of academic-lecture-style speech. *Frontiers in Psychology*, *9*. https://doi.org/10.3389/fpsyg.2018.01514

Shultz, A. A., Francis, A. L., & Llanos, F. (2012). Differential cue weighting in perception and production of consonant voicing. *Journal of the Acoustical Society of America*, *132*, EL95–EL101. https://doi.org/10.1121/1.4736711

So, W. C., Kita, S., & Goldin-Meadow, S. (2009). Using the hands to identify who does what to whom: Gesture and speech go hand-in-hand. *Cognitive Science*, *33*, 115–125. https://doi.org/10.1111/j.1551-6709.2008.01006.x

So, W. C., Sim Chen-Hui, C., & Low Wei-Shan, J. (2012). Mnemonic effect of iconic gesture and beat gesture in adults and children: Is meaning in gesture important for memory recall? *Language and Cognitive Processes*, *27*, 665–681. https://doi.org/10.1080/01690965.2011.573220

Spalek, K., Gotzner, N., & Wartenburger, I. (2014). Not only the apples: Focus sensitive particles improve memory for information-structural alternatives. *Journal of Memory and Language*, *70*, 68–84. https://doi.org/10.1016/j.jml.2013.09.001

Steedman, M. (2000). Information structure and the syntax-phonology interface. *Linguistic Inquiry*, *31*, 649–689. https://doi.org/10.1162/002438900554505

Sturt, P., Sanford, A. J., Stewart, A., & Dawydiak, E. (2004). Linguistic focus and good-enough representations: An application of the change-detection paradigm. *Psychonomic Bulletin & Review*, *11*, 882–888. https://doi.org/10.3758/BF03196716

Sueyoshi, A., & Hardison, D. M. (2005). The role of gestures and facial cues in second language listening comprehension. *Language Learning*, *55*, 661–699. https://doi.org/10.1111/j.0023-8333.2005.00320.x

Trude, A. M., & Brown-Schmidt, S. (2012). Talker-specific perceptual adaptation during online speech perception. *Language and Cognitive Processes*, *27*, 979–1001. https://doi.org/10.1080/01690965.2011.597153

Vilà-Giménez, I., Igualada, A., & Prieto, P. (2019). Observing storytellers who use rhythmic beat gestures improves children's narrative discourse performance. *Developmental Psychology*, *55*, 250–262. https://doi.org/10.1037/dev0000604

Wang, L., & Chu, M. (2013). The role of beat gesture and pitch accent in semantic processing: An ERP study. *Neuropsychologia*, *51*, 2847–2855. https://doi.org/10.1016/j.neuropsychologia.2013.09.027

Watson, D. G., Tanenhaus, M. K., & Gunlogson, C. A. (2008). Interpreting pitch accents in online comprehension: H* vs. L+ H*. *Cognitive Science*, *32*, 1232–1244. https://doi.org/10.1080/03640210802138755

Weber, A., Braun, B., & Crocker, M. W. (2006). Finding referents in time: Eye-tracking evidence for the role of contrastive accents. *Language and Speech*, *49*, 367–392. https://doi.org/10.1177/00238309060490030301