Is adults' ability to interpret iconicity shared between the spoken and gestural modalities?

Mingtong Li(Mingtong.li@warwick.ac.uk)
Department of Psychology, University of Warwick, UK

Suzanne Aussems (S.Aussems.1@warwick.ac.uk)
Department of Psychology, University of Warwick, UK

Sotaro Kita (S.Kita@warwick.ac.uk)
Department of Psychology, University of Warwick, UK

Abstract

Iconicity (the resemblance between form and meaning) exists in various modes of communication. This study investigated whether adults interpret iconicity in speech and gesture via a modality-independent ability. We tested 348 adult participants and assessed their ability to use iconic prosody and iconic gesture cues when interpreting novel verb meanings. We manipulated the rate of the spoken novel verbs (iconic prosody) and the rate of observed hand movements (iconic gestures) to be either fast or slow in two verb-action matching tasks. Adults could use these iconic speed cues to interpret novel verbs as referring to a fast or slow version of the same action. Adults showed similar performances in the two verb-action matching tasks: those who performed well in the iconic prosody task also performed well in the iconic gesture task. This positive correlation persisted even after controlling verbal working memory. Thus, we conclude that adults possess a modality-independent ability for interpreting iconicity.

Keywords: iconicity; prosody; co-speech gestures; speed; verb comprehension

Introduction

Iconicity is the resemblance between the form and meaning of a sign (Dingemanse et al., 2015). It can be observed in various modes of communication (e.g., Perniss & Vigliocco, 2014). For instance, in the gestural modality, physical size (i.e., large or small) can be depicted by the size of iconic gestures, providing additional information about the size of a referent (Beattie & Shovelton, 2002; Holler et al., 2009). In the spoken modality, the physical size of referents can also be conveyed by sound symbolism in speech. For instance, high front vowels, such as /i/, are associated with smaller physical size, whereas low back vowels, such as /a/, are associated with larger physical size (Sapir, 1929). Similarly, in English, vowels in size adjectives are suggestive of size in meaning, in particular for vowels /i/ for small and /a/ for large (Winter & Perlman, 2021). Furthermore, prosodic cues in the spoken modality can also convey information about the physical sizes of referents. For instance, participants who listened to sandwich advertisements presented in a low-pitched voice envisioned the sandwich being significantly larger compared to those who listened to the advertisement in a high-pitched voice (Lowe & Haws, 2017). In the examples of iconic representation of size, size information about the referent is iconically depicted in both modalities. Thus, the concept of iconicity transcends modalities. That is, it can be applied to different modes of communication, such as the visual and auditory modalities. However, it is unclear whether people also possess a general cognitive ability for interpreting iconicity in different modalities. Thus, this study aims to investigate this question.

So, how do people interpret iconicity in different modalities? We propose that people have a modality-independent cognitive mechanism for interpreting iconicity. Iconic mapping can occur both within and across modalities. For instance, when interpreting iconic speech cues, such as a slow speech rate with elongated vowel sounds depicting a slow movement (e.g., "A koala is *craaawling*."), individuals map acoustic features of speech to visual features of action, establishing a cross-modal mapping. When interpreting iconic gesture cues, such as hand gestures depicting the shape of an object (e.g., hands forming a circle to depict the shape of an apple), individuals map visual features of gesture to visual features of action, establishing a within-modal mapping. Both types of cross-modal and within-modal mappings can depict the same set of properties of referents. Thus, we propose that when interpreting iconicity, modality-specific information is first processed (e.g., acoustic features in speech; visual features in action), and then organized into modality-independent mental representations, focusing on the similarities between the key features across modalities (e.g., elongated vowels and the slowing crawling action manner). Subsequently, iconic cues are connected based on the similarity features, constituting a modality-independent process. Given the proposed modality-independent cognitive stage for iconicity detection processes, individuals who perform well in cross-modal iconic mappings (e.g., acoustic to visual) are expected to also excel in within-modal iconic mappings (e.g., visual to visual). Therefore, we predict that adults will demonstrate similar performances in both types of mappings.

The Current Study

The current study involved two verb-action matching tasks to evaluate how adult participants interpret iconic cues. The first task focused on adults' ability to interpret iconic cues in speech, while the second task focused on adults' ability to interpret iconic cues in gestures. During both tasks, participants were asked to watch a pair of videos featuring computer-modified slow and fast versions of the same action. They were also introduced to a novel verb by an actress in a third

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.*

video, who engaged with them and provided the iconic cues while saying (“Look! The boy/girl is *blicking*!”). In the first task, participants were presented with iconic speech cues embedded in the novel verbs (and no gesture cues). If the novel verb was produced at a fast rate, participants were expected to select the fast action as the referent, and if the novel verb was produced at a slow speech rate, participants were expected to select the slow action as the referent. In the second task, participants were presented with iconic co-speech gesture cues that accompanied the novel verbs (and no iconic speech cues). If the gestures were produced at a fast rate, participants were expected to select the fast action as the referent, and if the gestures were produced at a slow rate, participants were expected to select the slow action as the referent. In between these two tasks, participants also completed a forward-digit span task to assess their verbal working memory. Crucially, previous studies found no significant relationship between verbal working memory and iconic gesture comprehension (Nicoladis & Gagnon, 2020; Wu & Coulson, 2015). Thus, the verbal working memory task does not involve any substantial similarity mappings and it can be used as a control task in correlational analyses to establish the specificity of the relationship between the iconic speech task and the iconic gesture task. We did not use visuospatial working memory task as a control task, because visuospatial working memory capacity correlated with motor working memory capacity (Nicoladis & Gagnon, 2020), which contributed to iconic gesture comprehension (Wu & Coulson, 2015).

Predictions

Our first hypothesis is that participants should use iconic cues in both the spoken (H1a) and gestural modalities (H1b) for interpreting novel verbs. That is, in both verb-action matching tasks, they should select the fast action video when the speech-rate or gesture-rate is fast, and the slow action video when the speech-rate or gesture-rate is slow. We tested this by comparing their performance in each task against chance. More importantly, we hypothesize that participants' performances in these two tasks should be positively correlated (H2a). That is, participants who use iconic speech cues for interpreting novel verbs should also use iconic gesture cues for interpreting novel verbs and show similar task performances. Finally, we predict that this positive relationship should persist when controlling verbal working memory (H2b). If this latter hypothesis is confirmed, then this would rule out two alternative explanations. First, it may be the case that participants who generally possess good intellectual abilities will excel at both verb-action matching tasks. Second, it may be the case that participants who generally are more engaged in the study tasks excel in both verb-action matching tasks. Thus, we want to control these effects unrelated to our research question by including a verbal working memory task, which does not involve similarity mappings. If these predictions are confirmed, then this pattern of results would suggest that adults' interpretation of iconicity in the spoken and gestural modalities is at least partly governed by a modality-independent process.

Method

The hypotheses, methods, materials, and analyses of this study were pre-registered via the Open Science Framework (OSF) prior to data collection and can be accessed at `https://osf.io/jp6qn`. The raw data and analysis scripts are available via OSF at `https://osf.io/4mnqg`.

Design

This study included two verb-action matching tasks, in which the speed of speech and gesture cues was manipulated as either fast or slow. When combined with an iconic cue, the novel verb could be interpreted as referring to a fast or a slow action. Participants were asked to select one action in the two-alternative forced choice task as a referent for a novel verb. Each task included 12 trials and participants' responses were coded as target or distractor choices on each trial and then averaged across trials to create a proportion of target action choices for each verb-action matching task. Participants were also asked to complete 4 trials of a forward digit span task (adapted from Roembke & McMurray, 2021). Although the two verb-action matching tasks included an experimental manipulation of speed, the nature of this study is correlational. Specifically, this study explored relationships among three key variables: 1) the proportion of target action choices in the verb-action matching task with iconic speech cues, 2) the proportion of target action choices in the verb-action matching task with iconic gesture cues, and 3) the proportion of correct digits in the forward digit span task.

Participants

The final sample included 348 adult participants (169 males, 177 females, 2 non-binaries) recruited via the crowdsourcing platform Prolific Academic. The sample size was calculated using G*Power (Faul et al., 2009) with a significance level (α) of .05 and an expected power of 80%. The calculation was based on our pilot data ($n = 40$), which showed a positive Spearman's Rank Correlation between the performances in verb-action matching tasks with iconic speech cues and iconic gesture cues, $r = .15$, 95% CI [-.19, .45], $p = .376$. Participants in the final sample had a mean age of 43.84 years ($SD = 13.40$), ranging between 18 and 76 years. The ethnicity of the sample was as follows: 296 participants identified as White, 22 participants as Black (or Black British, Caribbean or African), 21 participants as Asian (or Asian British), 7 participants as Mixed (or Multiple Ethnic Groups), and 2 participants preferred not to disclose. All participants were located in the United Kingdom and reported English as their first language. To ensure high-quality data, participants were prescreened via Prolific Academic. Prescreening criteria included: 1) English as the first language and primary language, 2) no language-related disorders, 3) no literacy difficulties, 4) and a previous study payment approval rate of 99% or higher. Prescreening settings also included a representative gender-balanced sample. Participants were paid a rate of £9 per hour for their time. An additional 41 participants were tested but excluded from the analysis due to the following five reasons: 1) they showed a speed bias and selected exclusively on answers of the same speed in the verb-matching tasks ($n = 19$), 2) they failed at least two out of three attention checks ($n = 2$), 3) they reported writing down digits in the forward digit span task ($n = 6$), 4) they did not confirm English as their first language ($n = 12$), or 5) they failed several of the abovementioned checks ($n = 2$).

Materials

Speech Stimuli Speech stimuli were recorded on a MacBook Pro 2020 by a female native English speaker, who was blind to the study design and hypotheses. This voice actress recorded the carrier sentence, "Wow! Look at what is happening! The boy (or the girl) is *novel-verbing*. Which one is *novel-verbing*?", with 12 different novel verbs (*daxing, blicking, glabbing, howning, krading, larping, mipping, pilking, poffing, stumming, wepping*, and *yoofing*). These novel verbs follow the rules of the English language and are widely used in the verb learning literature (Aussems & Kita, 2021; Childers, 2011; Mumford & Kita, 2014). Then, the novel verbs were segmented using Praat (Version 6.1.16) speech analysis software (Boersma & Weenink, 2001) and the speech rate of each novel verb was computer-modified using Adobe Audition to create fast and slow versions.

Action Video Stimuli A set of 12 action video clips was selected from the GestuRe and Action Exemplar (GRACE) video database (Aussems et al., 2017, 2018). In the action video clips, an actor performed an unusual manner of locomotion to get from one location to another (i.e., from the left side to the right side of a scene), which looked like a funny manner of walking or jumping. The speed of selected locomotion manners was modulated using Adobe Premier Pro, to create a fast version and a slow version of each action video.

Gesture Video Stimuli A set of 12 iconic gesture templates for the selected action video clips were chosen from the same GRACE video database. The iconic mappings between action videos and corresponding iconic gestures have been normed and validated with adult participants (Aussems et al., 2017, 2018), as well as used in experiments with preschool-aged children (Aussems et al., 2021; Aussems & Kita, 2021). Following the validated iconic gesture templates, gesture video clips were recorded using a Canon R6 camera. The gesture videos were recorded for this study with an actress who was wearing a surgical mask. This was done to hide the mouth movements of the actress so that the different speech stimuli could be inserted in the exact same gesture video clips without creating a mismatch between her lip movements and what infants heard. In the gesture video clips, the actress performed an iconic gesture, depicting an action repeatedly. The gesture rate aligned with the moving rate of the manner of locomotion in the action video. The speed of gesture videos was then computer-modified in the same way as the speed of the action videos, to create a fast and a slow version of each gesture video.

Speech and Action Stimulus Pairings A total of 12 pairings were made, each consisting of an action and a novel verb. The same action was always paired with the same novel verb. A full list of the stimulus pairings can be found on the OSF project page: `https://osf.io/4mnqg`.

Software The study was run online via the crowdsourcing platform Prolific Academic. The study tasks were programmed in PsychoPy (Peirce et al., 2019) and run in Pavlovia (Bridges et al., 2020).

Procedure

All participants completed the same three cognitive tasks in the same order. First, participants completed a verb-action matching task with iconic speech cues (and no gesture cues). During this task, participants completed 12 two-alternative forced choice trials. In each trial, participants were presented with two video clips of the same actor who moved across the scene in a funny and unusual manner. One video showed a computer-modified slow version of the action, and the other video showed a computer-modified fast version of the same action. Simultaneously, participants were also presented with a third video clip, featuring an actress who sat on a chair wearing a face mask, accompanied by recorded speech introducing a new verb, "Wow! Look at what is happening! The boy (or the girl) is *novel-verbing*!". After the actress said the novel verb, participants were asked to select the target action that the actress was referring to, "Which one is *novel-verbing*?". The actress did not produce any gestures, but the speech rate of the novel verb was either fast or slow and could thus be mapped iconically onto the fast or slow action video. Participants' reaction times for each trial were also recorded for exploratory analysis.

Second, participants completed a forward digit span task adapted from Roembke & McMurray (2021). During this task, participants were required to recall sequences of 8 digits in four trials. In each trial, a participant watched a sequence of 8 digits presented in the center of their screen. Participants were asked to mentally memorize the digits in order of appearance. Each digit was displayed for 1 second and disappeared before the next digit appeared. Immediately after a full sequence of digits was presented, participants were asked to type in the digits in order of appearance using their keyboard. Afterwards, participants were asked whether they wrote down the digits or used any aids during the task.

Third, participants completed a verb-action matching task with iconic gesture cues (and no iconic speech cues). This task followed the same procedure as the verb-action matching task with iconic speech cues, except the actress now provided the iconic cue about the speed of the referent action (fast or slow) in gesture and not in speech. While she said the novel verb with a normal speech rate, she now also produced an iconic gesture at a fast or slow rate, which could thus be mapped iconically onto the fast or slow action video.

Before and after the three study tasks, three attention checks appeared where participants had to listen to an audio clip and type in the number of beeps they heard. Finally, participants answered demographic questions about their age, gender, ethnicity, and language background.

Counterbalancing and Randomization

The position of the target action was counterbalanced across trials. Half the time the target action appeared on the left side of the screen and half the time on the right side of the screen. The speed of the target actions was also counterbalanced across trials. Half the time the target actions were fast actions and half the time the target actions were slow actions.

Furthermore, whether the fast or slow version of an action was the target action was counterbalanced across two versions of the task. For a specific action, one version of the task included the fast version as the target action and the other version included the slow version.

The novel verbs and actions differed between the verb-action matching task with iconic speech cues and the verb-action matching task with iconic gesture cues. The verb-action pairings were always displayed in a pre-randomized, fixed order for all participants. Participants were randomly assigned to one version of the task, while ensuring that an equal number of participants completed each version.

Data Analysis

The proportions of target action choices for both verb-action matching tasks (with iconic speech or gesture cues) were created in the following way. For each trial, participants received a score of 1 when they chose the target action, and a score of 0 when they chose the distractor action. These scores were then summed across all trials and divided by the total number of test trials ($n = 12$) to create a proportion of target action choices per verb-action matching task. For instance, if a participant chose the target action in 9 out of 12 trials, the proportion of the target action choices was .75.

For the forward digit span task, we measured the proportion of correct digits in relation to the sequence of digits presented in the task. A digit entry was considered correct when both the digit itself and its position matched with the original sequence. Participants received a score of 1 for a correct digit, and a score of 0 for an incorrect digit. These scores were then summed across all digits of the same sequence and divided by the total number of digits in this sequence ($n = 8$) to create a proportion of correct digit entries for a given sequence. For instance, if a participant viewed a sequence of 8 digits as "412356785" and entered "412367855", the proportion of the correct digits was .625. This calculation considered only the first four digits '4123' and the last digit '5' as correct (i.e., both the digit and position had to match the original sequence of 8 digits). Finally, the proportions of correct digits across the forward digit span trials were summed and divided by the total number of trials ($n = 4$) to create an average proportion of correct digits.

To test H1a and H1b, two separate one-sample t-tests (two-tailed) were conducted to compare the average proportion of target choices in each verb-action matching task against chance (test value = .50). To test H2a, a correlation was conducted between the two average proportions of target action choices in the verb-action matching tasks. These proportions were also correlated with the proportion of correct digits across the forward digit span trials to provide a complete overview of the relationships between the three key variables. Finally, to test H2b, a partial correlation was conducted between the average proportions of target action choices in the verb-action matching tasks, while controlling the average proportion of correct digits across the forward digit span trials. We used Spearman Rank's correlation for the correlational analysis as proportions are not normally distributed by definition.

All analyses were conducted with R (version 4.2.2) and R Studio software (version 2023.06.2) for statistical analysis (R Core Team, 2022). The following R packages were used: *ggplot2* (Wickham, 2016) for data visualization, *rstatix* (Kassambara, 2021) for one-sample t-tests, and *ppcor* (Kim, 2015) for partial correlations.

Results

Chance Comparisons

Figure 1 shows participants' performances (in proportion) organized by verb-action matching task. The average proportion of target action choices of .58 ($SD = .16$) in the verb-action matching task with iconic speech cues was significantly above chance, $t(347) = 9.89, p < .001$, with 95% CIs [.56, .60]. The magnitude of this effect was moderate, Cohen's d = 0.53, 95% CIs $[0.41, 0.65]$. The average proportion of target action choices of .65 ($SD = .20$) in the verb-action matching task with iconic gesture cues was also significantly above chance, $t(347) = 13.85$, $p < .001$, with 95% CIs [.63, .67]. The magnitude of this effect was moderate, Cohen's $d = 0.74$, 95% CIs $[0.63, 0.87]$.

Correlations

The forward digit span task showed a good spread of performance ($M = .74$, $SD = .19$), which is appropriate for correlational analysis. Table 1 shows the correlations between the three key variables. There was a significant positive correlation between verb-action matching tasks with iconic speech cues and iconic gesture cues, $\rho(348) = .13$, $p = .016$, bootstrapped 95% CIs $[.03, .23]$, and a significant positive correlation between verb-action matching task with iconic speech cues and the forward digit span task performance, $\rho(348) = .12$, $p = .021$, bootstrapped 95% CIs $[.02, .23]$. No significant correlation was found between the verb-action matching task with iconic gesture cues and the forward digit span task performance. Figure 2 illustrates the relationship between the average proportions of target action choices in the verb-action matching tasks with iconic speech cues and iconic gesture cues.

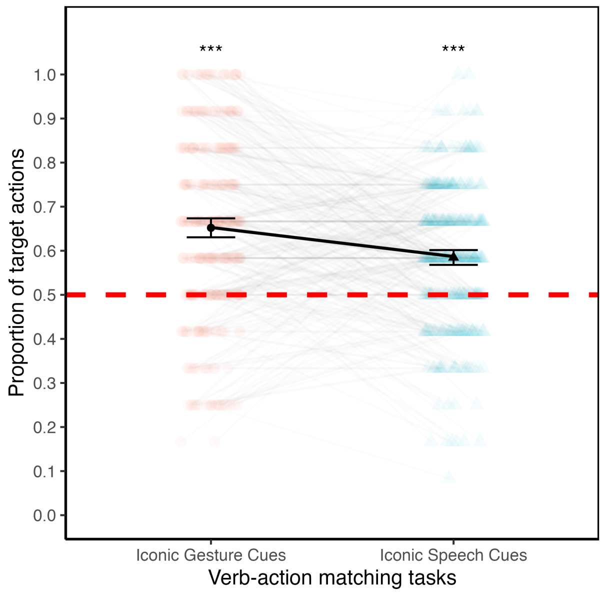

Figure 1: Performances (in proportion on the y-axis) organized by verb-action matching tasks (x-axis). The black shapes represent the means of all participants. Faded shapes represent individual performances, with light gray lines connecting the performances of the same individuals across the two verb-action matching tasks. Error bars represent 95% confidence intervals around the means. The dotted red horizontal line represents the chance level.

Table 1: Spearman's Rank Correlations between Verb-Action Matching Tasks and Forward Digit Span Task ($N = 348$)

	Task	1	2	3
1.	Verb-action matching (Speech)	–	$.13^*$ [.03, .23]	$.12^*$ [.12, .23]
2.	Verb-action matching (Gesture)		–	−.02 [−.13, .09]
3.	Forward digit span			–

Note. $^*p < .05$. Bootstrapped 95% CIs are reported in square brackets.

Partial Correlation

While controlling the average proportion of correct digits in the forward-digit span task, the significant positive relationship between proportions of target actions in the verb-action matching tasks with iconic speech cues and iconic gesture cues persists, $\rho(348) = .13$, $p = .014$, bootstrapped 95% CIs [.03, .24].

Exploratory Analysis

We examined whether the performance of verb-action matching tasks showed speed-accuracy trade-off effects. There were no significant correlations between reaction time and performances in both verb-action matching tasks with iconic speech cues ($\rho(348) = -.01$, $p = .065$, bootstrapped 95% CIs [−.21, .00]), and verb-action matching task with iconic gesture cues ($\rho(348) = .01$, $p = .876$, bootstrapped 95% CIs [−.09, .12]). Thus, there were no speed-accuracy trade-off effects found in the verb-action matching tasks.

Discussion

This study examined whether adults interpret iconicity in the spoken and gestural modalities via a modality-independent ability. There are three key findings. First, adults who inferred the meaning of novel verbs via iconicity selected the target action significantly above chance, when presented with iconic speech cues or iconic gesture cues that conveyed speed information about the action referent (i.e., speech rate or gesture rate is either fast or slow), confirming H1a and H1b. Thus, adults can glean information about verb referents from iconic cues, particularly speed-related iconic cues, in both the spoken and gestural modalities. Second, the performance in the verb-action matching task with iconic speech cues significantly positively correlated with the performance in the verb-action matching task with iconic gesture cues, confirming H2a. Thus, adults' ability to interpret iconicity appears to have a processing stage shared across the spoken and gestural modalities, in addition to modality-specific information processing. Third, this significant positive relationship persisted when controlling verbal working memory, confirming H2b. Thus, it is unlikely that our effect is due to the general intellectual ability or the general level of task engagement.

The crucial result is that, interpreting iconic speed information in the spoken modality is positively correlated with interpreting iconic speed information in the gestural modality. The former task is a cross-modal mapping, associating acoustic features (i.e., speech rate) with visual features (i.e., action speed), whereas the latter task is a within-modal mapping, associating visual features (i.e., gesture rate and action speed). While there could be individual differences in the ability of processing modality-specific information in speech and gestural modalities, this positive correlation suggests that there is a modality-independent cognitive mechanism where acoustic and visual features are mapped onto action representations based on similarity. That is, there is a modality-independent processing stage in the iconic mapping process.

The finding that adults can interpret iconic speed information conveyed by speech rate when inferring the meaning of a novel verb is consistent with previous studies on adults' comprehension of expressive speech cues (Kunihira, 1971; Shintel & Nusbaum, 2007). English-speaking adults who heard Japanese antonym pairs pronounced with an expressive voice

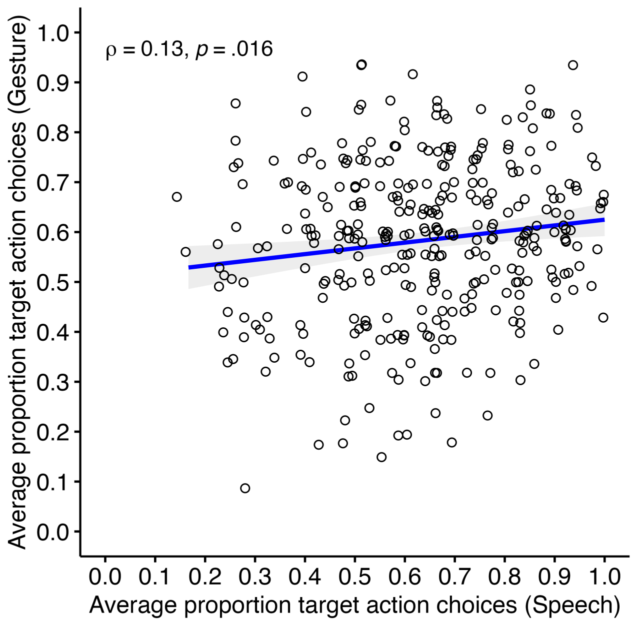

Figure 2: Performances of participants in the two verb-action matching tasks (iconic gesture cues vs. iconic speech cues). Jittered dots represent individual performances. The blue line represents the regression line and the grey area around the line represents bootstrapped 95% CIs. Spearman's Rank Correlation was significant, $\rho(348) = .13$, $p = .016$.

guessed the meaning of these words more accurately compared to those who heard with a monotone voice (Kunihira, 1971). In particular, the proportion of correct answers for the action verb anonym pair, *Aruku* (walk) and *Hashiru* (run), increased from .52 with a monotone voice to .70 with an expressive voice. This suggests that English-speaking adults can gain knowledge about novel verbs from expressive speech cues. Furthermore, English-speaking adults were sensitive to the speed information conveyed by speech rates when hearing sentences describe objects and integrated this speed information with objects when later being asked to recognize these pictures of these objects in either motion or static status (Shintel & Nusbaum, 2007). For instance, when they heard sentences like "The horse is brown" with a fast speech rate, they were more likely to correctly recall a picture of a moving horse (motion condition), compared to a picture of a standing horse (static condition). This further suggests that adults can interpret speed information conveyed by the speech rate of a sentence and implicitly integrate this information with its referent. Additionally, a developmental study by Hupp and Jungers (2013) suggested preschoolers can interpret iconic speech information conveyed by speech rate. Four- to six-year-old children and adults demonstrated the ability to accurately select a fast-moving star when they heard a sentence with a fast speech rate, and a slow-moving star when they heard a sentence with a slow speech rate.

In addition, the finding that adults can interpret iconic speed information conveyed by gesture rate when inferring the meaning of a novel verb is also consistent with previous studies. Adults can gain additional information from co-speech gestures, especially with iconic gestures, which are semantically correlated to the accompanying speech and depict a concrete referent (for a systematic meta-analysis, see Dargue et al., 2019; Hostetter, 2011). Adults learn more foreign words when taught while seeing iconic gestures compared to meaningless movements (Macedonia et al., 2011), and their semantic judgement about a verb was facilitated by an iconic gesture that was congruent with the verb meaning, as compared to an incongruent iconic gesture (Kelly et al., 2009).

Furthermore, our findings go beyond the previous research on the role of iconic gestures in verb comprehension. Previous research mainly focused on how adults benefited from iconic gestures that depicted the manner or the path of locomotion when learning new words (e.g., Macedonia et al., 2011). This study is the first to investigate whether adults interpret iconic gestures representing different aspects of an action, such as speed conveyed through gesture rates, and whether they use this information when comprehending the meaning of a novel verb.

Conclusions

Our study demonstrates preliminary evidence that at least some aspect of adults' ability to interpret iconicity, is shared between the spoken and gestural modalities. This suggests that there may be a modality-independent process that enables adults to recognize the resemblance between form and meaning. Future studies should investigate whether this modality-independent process is used when judging iconicity in modalities other than vision and audition, such as the tactile modality. Finally, it would also be interesting to investigate if the ability to interpret iconicity is shared across modalities in children, and if not, when this occurs developmentally.

Acknowledgement

We extend our sincere gratitude to all the participants who contributed to this study. Furthermore, we would like to acknowledge the support provided by the Psychology Departmental Fellowship at the University of Warwick.

References

Aussems, S., & Kita, S. (2021). Seeing iconic gesture promotes first- and second-order verb generalization in preschoolers. *Child Development*, *92*(1), 124–141. doi: 10.1111/cdev.13392

Aussems, S., Kwok, N., & Kita, S. (2017). Digital resource to support: "GestuRe and ACtion Exemplar (GRACE) Video database: Stimuli for Research on Manners of Human Locomotion and Iconic Gestures". [online]. University of Warwick, Department of Psychology. Available via: http://wrap.warwick.ac.uk/78493/.

Aussems, S., Kwok, N., & Kita, S. (2018). Gesture and action exemplar (grace) video database: stimuli for research on manners of human locomotion and iconic gestures. *Behavior Research Methods*, *50*(3), 1270–1284. doi: 10.3758/s13428-017-0942-2

Aussems, S., Mumford, K. H., & Kita, S. (2021). Prior experience with unlabeled actions promotes 3-year-old children's verb learning. *Journal of Experimental Psychology: General*. doi: 10.1037/xge0001071

Beattie, G., & Shovelton, H. (2002). An experimental investigation of some properties of individual iconic gestures that mediate their communicative power. *British Journal of Psychology*, *93*(2), 179–192. doi: 10.1348/000712602162526

Boersma, P., & Weenink, D. (2001). Praat, a system for doing phonetics by computer. *Glot International*, *5*(9/10), 341–345.

Bridges, D., Pitiot, A., MacAskill, M. R., & Peirce, J. W. (2020). The timing mega-study: comparing a range of experiment generators, both lab-based and online. *PeerJ*, *8*, e9414. (PMID: 33005482 PMCID: PMC7512138) doi: 10.7717/peerj.9414

Childers, J. B. (2011). Attention to multiple events helps two-and-a-half-year-olds extend new verbs. *First Language*, *31*(1), 3–22. doi: 10.1177/0142723710361825

Dargue, N., Sweller, N., & Jones, M. P. (2019, August). When our hands help us understand: A meta-analysis into the effects of gesture on comprehension. *Psychological Bulletin*, *145*(8), 765–784. doi: 10.1037/bul0000202

Dingemanse, M., Blasi, D. E., Lupyan, G., Christiansen, M. H., & Monaghan, P. (2015). Arbitrariness, iconicity, and systematicity in language. *Trends in Cognitive Sciences*, *19*(10), 603–615. doi: 10.1016/j.tics.2015.07.013

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using g*power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149–1160. doi: 10.3758/BRM.41.4.1149

Holler, J., Shovelton, H., & Beattie, G. (2009). Do iconic hand gestures really contribute to the communication of semantic information in a face-to-face context? *Journal of Nonverbal Behavior*, *33*(2), 73–88. doi: 10.1007/s10919-008-0063-9

Hostetter, A. B. (2011). When do gestures communicate? a meta-analysis. *Psychological Bulletin*, *137*(2), 297–315. doi: 10.1037/a0022128

Hupp, J. M., & Jungers, M. K. (2013). Beyond words: Comprehension and production of pragmatic prosody in adults and children. *Journal of Experimental Child Psychology*, *115*(3), 536–551. doi: 10.1016/j.jecp.2012.12.012

Kassambara, A. (2021). *rstatix: Pipe-friendly framework for basic statistical tests* (Tech. Rep.). Retrieved from `https://CRAN.R-project.org/package=rstatix`

Kelly, S. D., Özyürek, A., & Maris, E. (2009, December). Two sides of the same coin: Speech and gesture mutually interact to enhance comprehension. *Psychological Science*, *21*(2), 260–267. doi: 10.1177/0956797609357327

Kim, S. (2015). *ppcor: Partial and semi-partial (part) correlation* (Tech. Rep.). Retrieved from `https://CRAN.R-project.org/package=ppcor`

Kunihira, S. (1971). Effects of the expressive voice on phonetic symbolism. *Journal of Verbal Learning and Verbal Behavior*, *10*(4), 427–429. doi: 10.1016/S0022-5371(71)80042-7

Lowe, M. L., & Haws, K. L. (2017). Sounds big: The effects of acoustic pitch on product perceptions. *Journal of Marketing Research*, *54*(2), 331–346.

Macedonia, M., Müller, K., & Friederici, A. D. (2011, May). The impact of iconic gestures on foreign language word learning and its neural substrate. *Human Brain Mapping*, *32*(6), 982–998. doi: 10.1002/hbm.21084

Mumford, K. H., & Kita, S. (2014). Children use gesture to interpret novel verb meanings. *Child Development*, *85*(3), 1181–1189. doi: 10.1111/cdev.12188

Nicoladis, E., & Gagnon, R. (2020). Towards a reliable measure of motor working memory: revisiting wu and coulson's (2014) movement span task. *Royal Society Open Science*, *7*(9), 200237. doi: 10.1098/rsos.200237

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., … Lindeløv, J. K. (2019). Psychopy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203. doi: 10.3758/s13428-018-01193-y

Perniss, P., & Vigliocco, G. (2014). The bridge of iconicity: from a world of experience to the experience of language. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1651), 20130300. doi: 10.1098/rstb.2013.0300

R-Core-Team. (2022). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from `https://www.R-project.org/`

Roembke, T. C., & McMurray, B. (2021). Multiple components of statistical word learning are resource dependent: Evidence from a dual-task learning paradigm. *Memory Cognition*, *49*(5), 984–997. doi: 10.3758/s13421-021-01141-w

Sapir, E. (1929). A study in phonetic symbolism. *Journal of Experimental Psychology*, *12*(3), 225–239. doi: 10.1037/h0070931

Shintel, H., & Nusbaum, H. C. (2007). The sound of motion in spoken language: Visual information conveyed by acoustic properties of speech. *Cognition*, *105*(3), 681–690. doi: 10.1016/j.cognition.2006.11.005

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. Retrieved from `https://ggplot2.tidyverse.org`

Winter, B., & Perlman, M. (2021). Size sound symbolism in the english lexicon. *Glossa: a journal of general linguistics*, *6*(1). doi: 10.5334/gjgl.1646

Wu, Y. C., & Coulson, S. (2015). Iconic gestures facilitate discourse comprehension in individuals with superior immediate memory for body configurations. *Psychological Science*, *26*(11), 1717–1727. doi: 10.1177/0956797615597671