AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Experimental Psychology: General

ISSN: 0096-3445

2024, Vol. 153, No. 7, 1904–1919
https://doi.org/10.1037/xge0001588

Communicative Efficiency in Multimodal Language Directed at Children and Adults

Beata Grzyb[1], Stefan L. Frank[2], and Gabriella Vigliocco[1]
[1] Department of Experimental Psychology, University College London
[2] Centre for Language Studies, Radboud University

The ecology of human communication is face to face. In these contexts, speakers dynamically modify their communication across vocal (e.g., speaking rate) and gestural (e.g., cospeech gestures related in meaning to the content of speech) channels while speaking. What is the function of these adjustments? Here we ask whether speakers dynamically make these adjustments to increase communicative success, and decrease cognitive effort while speaking. We assess whether speakers modulate word durations and produce iconic (i.e., imagistically evoking properties of referents) gestures depending on the predictability of each word they utter. Predictability is operationalized as surprisal and computed from computational language models trained on corpora of child-directed, or adult-directed language. Using data from a novel corpus (Ecological Language Corpus) of naturalistic interactions between adult–child (aged 3–4), and adult–adult, we show that surprisal predicts speakers' multimodal adjustments and that some of these effects are modulated by whether the comprehender is a child or an adult. Thus, communicative efficiency applies generally across vocal and gestural communicative channels not being limited to structural properties of language or vocal modality.

Public Significance Statement
In everyday language use, whether speakers are addressing a child or an adult, they modulate their speech and use gestures. The function of these multimodal, nonlinguistic behaviors has been investigated separately in different subfields (linguistics, gesture studies, language acquisition, and language processing). Here we bring these fields together in asking whether communicative efficiency accounts for multimodal (vocal and gestural) adjustments across different audiences (adult–child and adult–adult). Using computational methods and analyses of child-directed and adult-directed language from a new conversational corpus, we found that adult speakers, talking to young children or adults, increase communicative efficiency by modulating word durations and producing gestures depending upon the predictability of to-be uttered words. These findings have implications for our understanding of language showing that behaviors, traditionally considered as nonlinguistic, are dynamically modulated by linguistic content. They have implications for education as they provide further evidence of how caregivers shape the learning environment.

Keywords: communicative efficiency, gesture, prosody, child-directed language, adult-directed language

Supplemental materials: https://doi.org/10.1037/xge0001588.supp

Throughout language evolution, the main medium for language use has been face to face, and still, in modern humans, children learn language in interaction with caregivers, and adults very often use language in face-to-face (in person or online) contexts. In these contexts, speakers dynamically modify nonlinguistic multimodal behaviors accompanying speech. For instance, speakers modulate their intonation (e.g.,

This article was published Online First June 6, 2024.
Michele Diaz served as action editor.
Gabriella Vigliocco https://orcid.org/0000-0002-7190-3659
The work reported in this article was funded by the European Research Council Advanced Grant (ECOLANG, 743035) and the Royal Society Wolfson Research Merit Award (WRM\R3\170016) presented to Gabriella Vigliocco. A preprint of this article has been made publicly available at PsyArXiv (https://psyarxiv.com/a9wt3). The authors declare no conflicts of interest. Data files, analysis scripts, and other materials pertaining to the study can be found at the project's OSF page. This study was not preregistered.

Beata Grzyb served as lead for data curation, formal analysis, and visualization. Stefan L. Frank served as lead for supervision and served in a supporting role for formal analysis, methodology, writing–original draft, and writing–review and editing. Gabriella Vigliocco served as lead for funding acquisition, project administration, resources, and supervision. Beata Grzyb and Gabriella Vigliocco contributed equally to conceptualization, writing–original draft, writing–review and editing, and methodology.
Correspondence concerning this article should be addressed to Gabriella Vigliocco, Department of Experimental Psychology, University College London, UCL Psychology, 26 Bedford Way, London WC1 6BT, United Kingdom. Email: g.vigliocco@ucl.ac.uk

changing their pitch or word duration) depending on whether they talk to other adults (adult-directed language, ADL; Aylett & Turk, 2006; Pluymaekers et al., 2005) or to children (child-directed language, CDL; Fernald & Simon, 1984; Fernald et al., 1989; Soderstrom, 2007). Speakers also gesture, producing for example iconic gestures that imagistically evoke visual features of objects and properties of actions (e.g., when talking about a recipe, the movement of the hands can act out whisking an omelet). These gestures are common in both ADL (Kita & Ozyurek, 2003; McNeill, 1992) and CDL (Flevares & Perry, 2001; Goldin-Meadow, 1999; Iverson et al., 1999; Zammit & Schafer, 2011). In general, these modifications appear to be robust across different cultures and languages (Fernald et al., 1989; Zammit & Schafer, 2011), and across genders (Fernald et al., 1989). In addition, they are also found in speech to nonnative adult speakers (Uther et al., 2007), as well as in speech toward pets (Ben-Aderet et al., 2017). But what is the function of these multimodal behaviors? Here, we investigate the idea that, across ADL and CDL, they support efficient communication.

Multimodal Communicative Efficiency

The idea that communication is efficient, namely that speakers and listeners try to minimize costs associated with communication while maximizing the benefits (mutual understanding) has a long history in cognitive science (see Levshina, 2022 for a review) and has been formalized within information theory (Shannon, 1948). Here, successful communication is achieved when the received message is equal or only slightly different from the source message. To increase chances of communicative success, the speaker can, for example, speak more clearly, use longer sentences, and so on. However, any of these come with associated costs such as greater articulatory effort, longer time, and so on. Communication is then most efficient when successful communication is achieved with minimal effort on average by the sender (speaker) and receiver (comprehender). Thus, communicative efficiency is a cost-to-benefit ratio (Gibson et al., 2019; Jaeger & Tily, 2011).

An important type of cost for the speaker is related to conveying messages that are cognitively less accessible and that for the listener would be unexpected, atypical, or unpredictable (Levshina, 2022). Thus, an efficient communication system would trade-off between reducing effort in communicating what is more accessible while increasing effort in communicating what is less accessible. Indeed, there is evidence that speakers tend to use shorter words to communicate more accessible referents (Mahowald et al., 2013; Meylan & Griffiths, 2021; Piantadosi et al., 2011; Zipf, 1949), shorter morpho-syntax (Haspelmath, 2021; Kurumada & Jaeger, 2015; Levy & Jaeger, 2007) and phonology (Hall et al., 2018; Priva, 2008; Seyfarth, 2014).

As language is for the most part learned and used in face-to-face contexts in which information conveyed by words is ubiquitously accompanied by information from other vocal (prosodic) and gestural behaviors, communicative efficiency can affect not only structural properties of language (e.g., word length or syntactic structure) but, more generally, multimodal communication across vocal and gestural channels, including prosodic modulations and gestures. Here, we examine the extent to which multimodal behaviors such as dynamic online adjustments of word duration and the probability of producing iconic gestures are associated with how predictable words are in their immediate linguistic context.

According to the communicative efficiency proposal, if the speaker believes a word to be more predictable to the listener, the speaker can speak faster and avoid producing gestures, thus reducing the speaker's effort required, without compromising communicative success. However, if a word is considered less predictable, speakers may speak slower and produce gestures to increase the likelihood of communicative success even if this implies greater speaker effort.

Predictability further depends on the characteristics of a comprehender (audience design) and therefore it can differ for a young child and for an adult based on their different experiences with language. A second question is therefore whether speakers do adjust their multimodal behaviors in a communicative efficient manner both in CDL and ADL, thus taking into account predictability from the perspective of a child or an adult.

Crucially, considering multimodal language also allows us to assess whether adjustments related to predictability can be observed in both the vocal and gestural channels and, moreover, to what extent they are dynamically modulated in an (in)dependent way in CDL and ADL. Dependency between adjustments across channels would be expected if speakers adjust in one channel (to increase communicative success) but not in the other (to reduce effort). It could also arise in CDL if caregivers prioritize communicative success over effort and therefore tend to adjust both channels at the same time. Below we review previous relevant work concerning prosodic and gestural adjustments.

Prosodic Modulations

Speakers modulate their speech in different ways: They change loudness, pitch, and speaking rate, and all these different modulations affect comprehension (see Cutler, 1996 for a review). Most previous studies concerned with the communicative efficiency of prosodic modulations have focused on modulations of speaking rate and word duration. Speakers can modulate their speaking rate to overcome production difficulties (i.e., to increase production ease) while maintaining fluent speech. For example, less frequent and longer words are more difficult to retrieve and encode (Arnold, 2008; Gahl et al., 2012; Watson et al., 2008), therefore, reducing speaking rate provides more time for speakers to complete these processes without compromising fluency. However, speakers can also reduce their speaking rate to lengthen words that are less predictable to the comprehender (Aylett & Turk, 2006; Florian Jaeger, 2010). These modulations can be explained in terms of communicative efficiency: Predictable words carry less information and therefore can be shortened without risk of information loss; whereas less predictable words carry more information that could be lost in noise, jeopardizing communicative success. Thus, speakers could hyperarticulate less predictable words to improve comprehenders' chances of successfully processing them (Brennan & Clark, 1996; Clark & Fox Tree, 2002; Grice, 1989). Likewise, speakers could shorten more predictable words, as these are easier to process, thus lessening their articulatory effort. Several studies have shown a link between words' predictability in context and word duration in adult-to-adult conversation, showing that more predictable words have shorter durations than less predictable words in line with communicative efficiency (Demberg et al., 2012; Seyfarth, 2014).

Prosodic modulations, including a slower speaking rate, are a hallmark of CDL (Fernald & Simon, 1984; Fernald et al., 1989; Soderstrom, 2007). These have been argued to attract children's attention (Segal & Newman, 2015), communicate emotion and attitudes between the caregiver and the child (Fernald, 1992; Fernald et al., 1989), and facilitate children's speech perception and word comprehension (Cooper & Aslin, 1990; Fernald, 2000; Kuhl et al., 1997;

Stern et al., 1983). They have also been shown to support word learning (Cristia, 2013; Estes & Hurley, 2013; Zangl & Mills, 2007).

While these previous CDL studies indicate that caregivers adjust their speaking rate to support their children's learning and processing, they do not directly assess whether these adjustments subserve communicative efficiency along the same lines as has been suggested for ADL. In other words, do caregivers modulate their speaking rate online according to word predictability for their children? In a corpus-based study, Pate and Goldwater (2015) assessed if speakers (talking to their infant or to other adults) were more likely to lengthen words that were less predictable to their addressees and shorten words that were more predictable. Different measures of predictability were used, including a measure based on simple word frequency in the corpus and one taking into account the immediately preceding linguistic context. It was found that word duration was significantly correlated with word frequency in both ADL and CDL, however, only in ADL a significant correlation with the measure of predictability based on the preceding context was found. As the effects of word frequency on word duration may be accounted for in terms of production-internal mechanisms (Barry et al., 2001; Gerhand & Barry, 1999), these results suggest that communicative efficiency may not underscore prosodic adjustments in CDL. However, very different corpora were used in the analyses of CDL and ADL, therefore making the comparison more difficult. A more recent study investigating Swedish adults talking to their 2- to 33-month-old infants reported an effect of surprisal on articulatory rate (Sjons et al., 2017). However, frequency or any other lexical or sentence-level variable known to affect speaking rates was not controlled in this study, making its interpretation difficult. Thus, while there is clear evidence that adult speakers modulate their speaking rates when talking to another adult as a function of predictability, whether this is also the case for adults talking to their children is unclear.

Gestures

When people speak, they gesture. Just like prosodic modulations, people produce different types of cospeech gestures. They produce referential gestures, such as iconic gestures that imagistically evoke properties of referents in the speech, and points that index referents in the speech. They also produce nonreferential gestures, such as beat (rhythmic movements of the hand) and other pragmatic gestures (e.g., moving the hand toward an addressee), that do not bear a semantic relationship with the content of speech (see Kendon, 2004; McNeill, 1992; Vilà-Giménez & Prieto, 2021, among many others).

While any of these gesture types can potentially support communicative success, iconic gestures (e.g., drawing a wiggly line with the index finger while talking about someone drunk walking) provide an ideal testbed for exploring the association between predictability and gesture production. Iconic gestures can support the speaker by decreasing cognitive load (Goldin-Meadow, 1999), priming conceptual information (Krauss et al., 2000), activating or maintaining spatial information (De Ruiter, 2000; Friedman, 1977), and preparing and structuring information for speaking (Kita et al., 2017). Iconic gestures can also be used by comprehenders as is apparent from the fact that speakers gesture more when the listener can see them (Hoetjes et al., 2015) and in situations when speakers believe that gesture might benefit comprehenders (Alibali et al., 2001; Kelly et al., 2011); and that speakers gesture less when comprehenders are familiar with the content (Galati & Brennan, 2014; Jacobs & Garnham, 2007). Iconic gestures have also been shown to help comprehenders disambiguate the meaning of words (e.g., the word ball presented with a gesture congruent with either the dance or toy meaning; Holle & Gunter, 2007), provide additional semantic information about upcoming words (Hostetter, 2011), and prime subsequent words and semantic concepts (Wu & Coulson, 2007; Yap et al., 2011). Thus, iconic gestures can support communicative success.

Two previous studies have investigated the association between word predictability and presence of iconic gestures (Hintz et al., 2023; Zhang et al., 2021). Both studies show that iconic gestures support the processing of the words they are semantically associated with. For example, Zhang et al. (2021) asked whether the N400—an electrophysiological measure that increases when processing words that are less predictable given their preceding linguistic context (Frank et al., 2015)—is modulated by observing iconic gestures. Subjects watched videos of a speaker narrating short passages while spontaneously gesturing. A significant reduction in N400 amplitude was found when less predictable words were accompanied by iconic gestures indicating that gestures supported the processing of these words. Given that gesture production tends to precede by 300–500 ms the production of the corresponding word, it makes sense that gestures increase the predictability of upcoming words (Donnellan et al., 2022; Ter Bekke et al., 2020).

These findings underscore the plausibility of assuming that iconic gestures may be more likely associated with less predictable words. More directly related to this claim, Beattie and Shovelton (2000) in an analysis of 12 utterances, were the first to report that speakers were indeed more likely to produce an iconic gesture when the lexical affiliate of the gesture (the lexical affiliate of a gesture is the word most related in meaning to the meaning expressed in the gesture) had a lower transitional probability (established using a cloze procedure), as predicted on the basis of the communicative efficiency account. However given the relatively small set of observations, it is impossible to know whether the words accompanied by a gesture were simply more gesturable because concrete and associated with visual and action semantic features more easily depicted in gestures.

The gestures parents use in interactions with their children have been shown to influence children's gesture development (Acredolo & Goodwyn, 1988; Özçalışkan & Goldin-Meadow, 2005), which in turn correlates with vocabulary development (Rowe et al., 2008). Children begin to produce iconic gestures relatively late (around 26 months; Goldin-Meadow, 1999). For this reason, many studies investigating caregiver's gestural production and its impact on learning have focused on other gestures such as pointing (Rowe et al., 2008). However, there is evidence that children as young as 18 months can successfully map iconic gestures onto novel object referents (Namy et al., 2004) and that iconic gestures facilitate the learning of novel words compared with arbitrary gestures in children between 2 and 4 years old (Goodrich & Hudson Kam, 2009; Vogt & Kauschke, 2017), as well as facilitating children's memory representation of events (Aussems & Kita, 2019). No previous study has asked whether caregivers adjust their gesture production on the fly in line with word predictability for the child as predicted by a communicative efficiency account.

The Present Study: Efficiency in Multimodal Communication

Both word lengthening and gesture production can be used by speakers to support communicative success. For words that are

more predictable, given the prior context, production effort can be reduced by shortening the word and/or avoiding to produce a gesture. For less predictable words, communicative success can be enhanced by lengthening the word and/or producing an associated gesture.

While communicative efficiency should account for production of multimodal adjustments regardless of whether the addressee is an adult or a child (and therefore whether the comprehender is primarily processing or learning language), there also are reasons to believe that "audience design" (Bell, 1984) may modulate effects. Considering word durations, while there is clear evidence that adults adjust their speech to other adults in a communicative efficient manner, for CDL previous evidence is mixed. As mentioned above, modulations of speaking rate (along with other prosodic modulations) in CDL have been argued to serve a number of different functions including attracting children's attention (Segal & Newman, 2015) and affective engagement with the caregiver (Fernald, 1992; Fernald et al., 1989) which are unrelated to predictability (although might be argued to also support communicative success). Note here that the age of the child may also modulate the effects with CDL prosodic modulations to infants (as in the studies above) serving attention and engagement processes, but as their vocabulary grows, serving communicative efficiency (more). There is evidence that 2-year-old children are sensitive to predictability in language (Mani & Huettig, 2012). Regarding gestures, while there is initial evidence that ADL shows effects compatible with communicative efficiency (Beattie & Shovelton, 2000) this finding needs to be replicated in a larger corpus. Concerning, CDL, previous work indicates that by age 3, children can produce and understand iconic gestures (Özçalışkan & Goldin-Meadow, 2005), but it is simply unknown whether caregivers produce these gestures in a communicative efficient manner, namely more often for words that are more surprising for their children.

Investigating adjustments across two modalities (vocal and gestural) provides us a unique opportunity to assess whether speakers modulate word duration and produce gestures in a dependent manner. In particular, if speakers are communicatively efficient, we might observe that they choose a channel (e.g., gesture) in which to modulate but also refrain from modulating in the other channel (e.g., prosody) to balance out between effort and success. This can be, however, modulated by audience design: if caregivers prioritize communicative success at the expense of increased effort, we should observe that both modulations are likely to be carried out at the same time in CDL but not in ADL.

We focus on word durations rather than other prosodic modulations in continuity with previous studies and for methodological reasons. For word durations we can obtain fully independent and objective baseline measures, for example from text-to-speech synthesizers, while for pitch (mean F0 or other measures) it is more debatable what a suitable baseline is (e.g., mean utterance F0 could be suitable for ADL, but it is unclear whether it would be suitable for CDL). Finally, intensity is highly dependent upon the position of the microphone during recording which was not kept constant in the corpus used here. We focus on iconic gestures because: (a) in contrast to points that require a physically present referent they can be produced for any (gesturable) content; (b) in contrast to non-referential gestures (like beats or pragmatic gestures), they are documented in the literature for both CDL as well as ADL, and moreover they are less related to prosody.

We model predictability for a comprehender—child or adult—and we assess whether predictability correlates with speakers' prosodic modulation and production of gestures above and beyond factors related to production difficulty which are known to also correlate with prosodic modulation and gesture production, such as age of acquisition (AoA) and word frequency (Humphreys et al., 1988; Jescheniak & Levelt, 1994; Oldfield & Wingfield, 1965). We use a new corpus of multimodal language (ecological language corpus [ECOLANG]) where a speaker talks to their child (3–4 years old) or to a familiar adult about objects provided by the researchers. In line with previous studies investigating communicative efficiency, we operationalize words' predictability in terms of surprisal, an information-theoretic measure of how unexpected a word is given the prior linguistic context (Levy, 2008; Shannon, 1948). Surprisal has been shown to reliably predict word reading times and electrophysiological correlates (N400) of processing difficulty in sentences (Frank et al., 2015; Goodkind & Bicknell, 2018; Lowder et al., 2018; Smith & Levy, 2013).

Here, surprisal measures were obtained from recurrent neural network (RNN) language models trained on corpora of CDL or ADL. We used RNNs because they have been shown to successfully account for psycholinguistic phenomena (Frank et al., 2016; Gulordava et al., 2019; van Schijndel & Linzen, 2021). Moreover, in contrast to other models such as transformers, they do not require massive corpora for training, which do not exist for CDL. Different corpora were used for CDL and ADL because the predictability of a word in context differs between child- and adult-directed language, as CDL is simplified, more repetitive, and has shorter sentences (Fernald, 1992). We test: (1) whether our measures of surprisal predict word durations and the probability of producing a gesture of gestures across CDL and ADL, (2) whether these adjustments are carried out independently in each channel (gesture, prosody) and finally (3) whether the effects are similar in CDL and ADL.

Method

Language Models

We implemented the language models as recurrent neural networks using gated recurrent units (GRUs; Cho et al., 2014 as GRUs provide a better language model than the more traditional simple recurrent network (Elman, 1990). When they are trained on relatively small text corpora like the ones we use here, GRUs account for human processing data to a similar extent as the more modern transformer architecture that current large language models are based on (Merkx & Frank, 2021). We used GRU networks with one input and one output unit for each word type, a 500-unit recurrent layer, a 400-unit hidden layer in between the input and recurrent layers, and a 400-unit hidden layer in between the recurrent and output layers (Aurnhammer & Frank, 2019). Hidden- and recurrent-layer units had tanh activation function, meaning that their activations are constrained to range between −1 and +1. The output layer uses the softmax activation function so that each unit's output activation is an estimate of the occurrence probability of the corresponding word given the utterance so far. This is turned into a surprisal value (Hale, 2001; Levy, 2008) by taking the negative logarithm:

$$\text{surprisal}(w_t) = -\log P(w_t|w_1, \ldots, w_{t-1}), \quad (1)$$

where $w_1, \ldots, w_{t-1}$ is the sequence of words in the utterance so far, w_t is the current word, and its surprisal formalizes the extent to which its occurrence was unexpected to the language model. During model

training and testing, the recurrent unit activations are reset to zero before each new utterance so that each utterance is processed independently from the previous one.

If the test utterances contain words that were not present in the corpus used for training, these words do not have corresponding input and output nodes in the network and cannot be assigned a surprisal value. To avoid this, we added to the training corpora a list of word types that were in the test utterances but not in the training data. Because these words are presented only once and without context, the network learns only about their existence but not about their use.

CDL-Based Language Model

As training data, we used caregiver's speech production extracted from 42 data sets of child-directed speech from the CHILDES corpus (the corpora are listed on our Open Science Framework [OSF] project's page https://osf.io/ugs28/). We selected only data sets that focus on the dyadic interaction between a caregiver and a child aged anywhere between 6 months up to (and including) 4 years and the language used by the caregiver was either British English or American English. We focused on this particular age range because the age of the children in our test set is 3–4 years. The data were preprocessed: the punctuation removed and all words converted to lowercase. In total, our child-directed language training set consisted of 2,364,255 utterances (45,963 word types). The mean utterance length was 4.46 words ($SD = 3.65$), and 21.57% of all sentences were single-word utterances (i.e., "isolations").

ADL-Based Language Model

As training data, we used the Spoken British National Corpus 2014 (Spoken BNC2014; Love et al., 2017; Figure 1C). This corpus contains transcripts of the audio recordings of face-to-face conversations of British English native speakers recorded in informal contexts. Conversations between two and three people constitute 71% of all conversations. As transcripts are in a free-flowing style (i.e., no punctuation), a deep neural network DeepSegment[1] was first used to split passages into utterances. All words were converted to lowercase. The resulting training set consisted of 1,566,534 utterances (67,797 word types). The mean utterance length was 6.58 words ($SD = 6.50$), and 16.36% of all utterances were isolations.

ECOLANG Multimodal Corpus

Speech and iconic gesture test data were taken from a new multimodal corpus of seminaturalistic interactions between a caregiver and a child (Ecolang_Child) or between two familiar adults (Ecolang_Adult; Gu et al., n.d.). The interaction sessions were carried out at the family home (for Ecolang_Child) or in the lab (for Ecolang_Adult). The participants were seated at a table with the speaker and the addressee sitting at 90° from each other. The interaction was video-recorded with two video cameras; in addition, the speaker also wore a lapel microphone.

In each interaction, speakers were asked to talk as they typically do with their addressees about objects taken from four categories: foods, musical instruments, animals, and tools. Participants talked about sets of six objects (three known and three unknown to the comprehender) taken from one category at a time (24 objects in total). Objects (toys) for Ecolang_Child were selected for each child from a larger set of about 20 toy items per category, each of which was used for roughly equal numbers of participants. Familiarity (known/unknown) of the toys was established by giving caregivers a list of our toys before the session and asking them to indicate which were known or unknown to their child. For the Ecolang_Adult corpus, two sets of 15 objects were prepared, one with objects generally unknown to the participants (e.g., strigil, axolotl, kalimba, rambutan) and one with objects most likely to be known to the participants, although generally uncommon (e.g., penguin, accordion, kiwi, chopsticks). As for Ecolang_Child, each dyad was presented with a total of 24 objects taken from the 30 prepared for the study. The participant who signed up for the study was assigned to the speaker's role and was asked to come to the lab with a friend/partner (the addressee). Before the interaction session, the speaker was taught about the unknown objects so that they could then talk about them to the addressee. Dyads talked about each set of objects for about 4–5 min and the entire recording session took 45–60 min. Note that for each set of objects, the interaction was repeated twice, once with the objects present and once with the (same) objects absent (counterbalanced). This manipulation was introduced to investigate multimodal differences in situated and displaced language which are not relevant to the present study but are controlled for in the analyses. To date, only the speakers' (i.e., not the addressees') multimodal behaviors have been annotated and therefore all analyses reported below only include speech and gestures from the speaker.

Word Duration Data

CDL. We extracted speech data from 33 (available at the time) caregiver–child dyads. Caregivers' age was between 29 and 48 years (mean age = 38.45). Children were aged between 36 and 52 months (median age = 41 months). The language used was British ($N = 30$) or American ($N = 3$) English. Caregivers' speech was transcribed manually in Praat (Boersma & Weenink, 1992–2022). The Montreal Forced Aligner (McAuliffe et al., 2017) tool or the Munich Automatic Segmentation System (MAUS[2]) was used to segment and align orthographically transcribed speech to the corresponding audio at the word level. The automatic alignment of word boundaries was then corrected manually in Praat. Overall, the mean number of utterances per participant was 772.48 ($SD = 104.26$). Across all dyads, single-word utterances constituted on average 21.86% ($SD = 5.08\%$) of all utterances, while on average the length of utterances was 4.65 words ($SD = 0.53$).

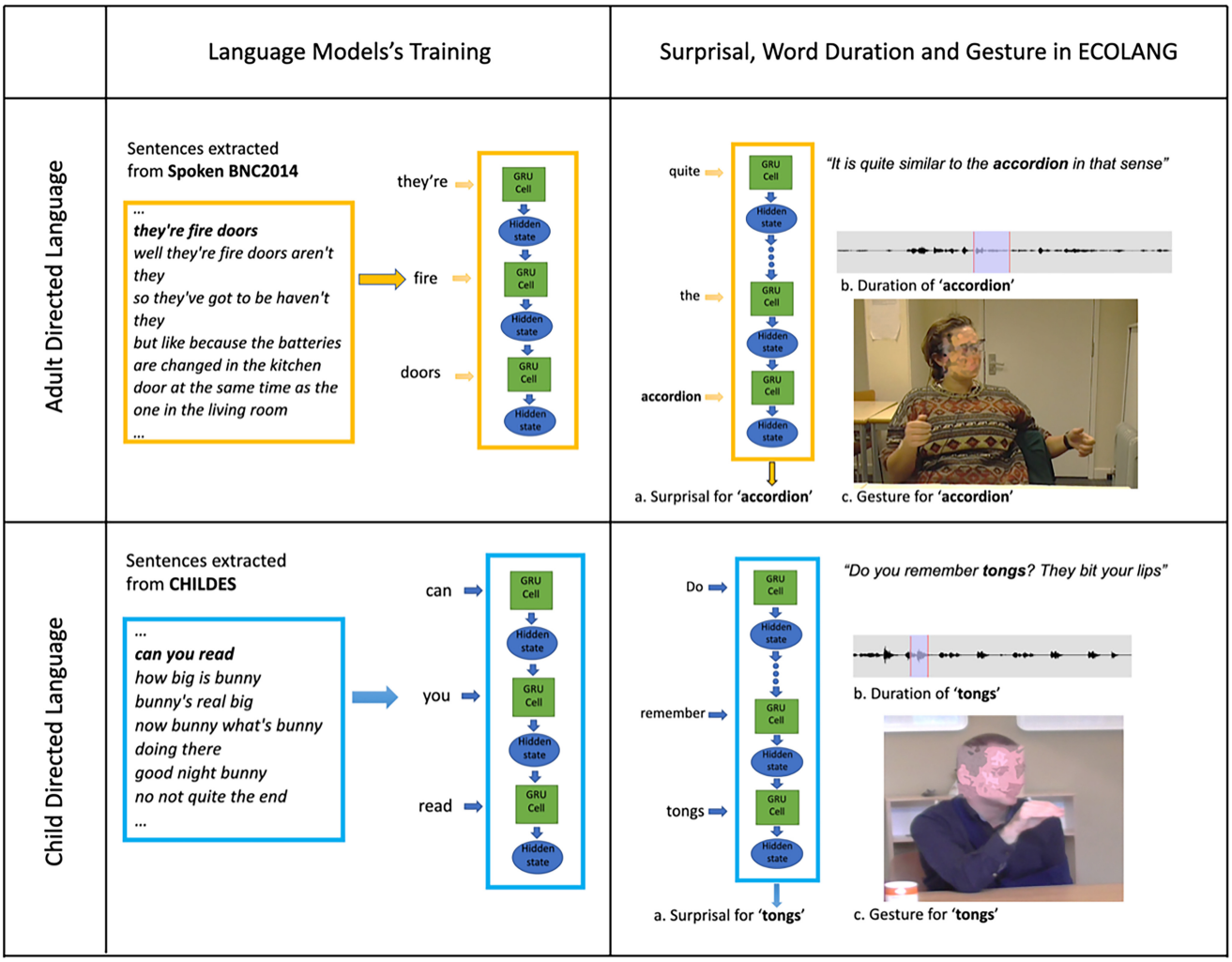
ADL. We used speech extracted from 30 (available at the time) adult dyads, and, as for the CDL corpus, only one of the partners' (the speaker, see below) communicative behaviors was used here. The mean ages of speakers (19 females and 14 males) were 24.24 years, range 18–43, $SD = 6.15$, and of addressees (17 females and 16 males) were 24.76 years, range 18–47, $SD = 6.29$. The language used was British ($N = 26$) or American ($N = 4$) English. Data were transcribed automatically using the TEMI system and then corrected manually in Praat. MAUS was used to segment and align orthographically transcribed speech to the corresponding audio at the word level. The automatic alignment of word boundaries was then corrected manually in Praat. Overall, the mean number of utterances per participant was 817.80 ($SD = 153$). Finally, 23.43% ($SD = 8.97\%$) of all utterances were single words, and on average utterances contained 5.71 words ($SD = 1.07$).

[1] https://github.com/bedapudi6788/deepsegment.

[2] https://www.en.phonetik.uni-muenchen.de/research/current_projects/maus_en.html.

Figure 1

Overview of the Methods



Note. Left panel: Training of ADL- and CDL-based language models implemented as GRU neural networks. Right panel: Extracting: (a) surprisal measures, (b) word durations, and (c) gestures from Ecolang_Child and Ecolang_Adult multimodal corpus for the words "accordion" and "tongs." All words of the utterance that precede a lexical affiliate (a word closely related to gesture in meaning) are passed through the language model, which calculates the surprisal of the lexical affiliate. ECOLANG examples are adapted from "The ECOLANG Multimodal Corpus of Adult–Child and Adult–Adult Conversation," by Y. Gu, E. Donnellan, B. Grzyb, G. Brekelmans, M. Murgiano, R. Brieke, P. Perniss, and G. Vigliocco (n.d.). BNC2014 sentences are adapted from "The Spoken BNC2014: Designing and Building a Spoken Corpus of Everyday Conversations," by R. Love, C. Dembry, A. Hardie, V. Brezina, and T. McEnery, 2017, International Journal of Corpus Linguistics, *22*(3), pp. 319–344 (https://www.jbe-platform.com/content/journals/10.1075/ijcl.22.3.02lov). Copyright 2017 by John Benjamins Publishing Company. Adapted with permission. CHILDES sentences are taken from the CHILDES corpora (https://childes.talkbank.org/access/). In the public domain. See the online article for the color version of this figure.

Iconic Gestures and Lexical Affiliates Data

We coded the use of iconic gestures that represent the semantic properties of referents. This may be through depicting the shape or size of an object (e.g., hands moving apart to represent the long legs of the flamingo) or how the object is manipulated (e.g., a hammering gesture in which the shape of the hand represents how a hand would hold a hammer). Gestures were coded manually in ELAN (Lausberg & Sloetjes, 2009), along with other gestures, in the corpus. Gesture coding was carried out manually by expert coding members of the team (see the project's OSF page for coding instructions: https://osf.io/ugs28/).

In addition, gestures' lexical affiliates were identified. Lexical affiliates were defined as the words that correspond most closely to an iconic gesture in meaning (e.g., "long" in "the flamingo has long legs," said while moving the hands apart to represent the length of the legs; Beattie & Shovelton, 2000; Hadar & Butterworth, 2009). Previous research has shown that these tend to be produced in close temporal proximity to the gestures (Donnellan et al., 2022). Lexical affiliates were not always single words. Function words (e.g., determiners, auxiliaries, prepositions, etc.) were not considered as lexical affiliate, except all those cases in which they were the most semantically appropriate word (e.g., in a gesture upward time-locked with

the utterance "up the stairs" the lexical affiliate would be "stairs" if the fingers are moving as the legs of a person; however, it would be "up" if it is a simple directional movement). Note also that it was not always possible to find a lexical affiliate for a gesture, as some gestures may convey information about utterance-level features (McNeill, 1992). These gestures were excluded from the analyses.

Because not all words are equally gesturable (e.g., concrete words are more gesturable than abstract ones), it would not be meaningful to simply compare words with and without gestures. We first identified all the gesturable word types. These are words that occurred at least once as lexical affiliate in the corpus. Then, for each participant, we extracted all tokens of the gesturable words along with their linguistic context, to compare the same words when produced with and without gestures in different contexts (and therefore with different surprisals).

CDL. We used data from 26 caregivers available at the time from the Ecolang_Child corpus and, for each caregiver, we coded iconic gestures and their lexical affiliates. There were 853 different gesturable word types. Overall, the mean number of gesturable word tokens per participant was 257.81 ($SD = 33.81$), and the mean number of lexical affiliates per participant was 70.73 ($SD = 43.15$; note that one gesture could refer to more than one word token).

ADL. We used data from 26 speakers available at the time from the Ecolang_Adult corpus, and for each speaker, we coded the iconic gestures and their lexical affiliates. As for the CDL data, we prepared a list of gesturable word types ($N = 1{,}276$). In total, the mean number of gesturable word tokens per participant was 301.77 ($SD = 49.52$), while the mean number of lexical affiliates per participant was 114.04 ($SD = 54.79$).

Gesture and Lexical Affiliate Coding Reliability

Ten percent of the total duration for the interaction was randomly selected for each participant for reliability coding. For gesture reliability, we analyzed the reliability of our coding scheme on data from these annotations by computing Spearman's correlation between main and reliability coders in the number of iconic gestures coded for the ADL and CDL corpus separately. Agreement between the coders on the number of representational gestures produced was high (CDL, $r = .823$, ADL, $r = .820$). For lexical affiliate reliability, we considered whether main and reliability coders agreed on the same word as lexical affiliate for a given gesture (1 = *they both agreed on the word, or both agreed that there was no lexical affiliate*; 0 = *they disagree on the word, or disagree whether there is a lexical affiliate*). For ADL, coders agreed 92% of the times (Cohen's $\kappa = .64$); for CDL, they agreed 89% of the times ($\kappa = .58$).[3]

Data Analysis

Analysis 1: Do Speakers Dynamically Produce Multimodal Adjustments Based on Surprisal in Both CDL and ADL?

Word Durations. We fitted a linear mixed-effects regression model using the R package lme4 on pooled child-directed and adult-directed language data. The dependent variable was observed log-transformed word duration in milliseconds. Here, a positive regression coefficient for surprisal indicates a longer duration for more surprising words.

The model included the following fixed effects of interest: group (CDL vs. ADL) coded with sum contrast scaled to values of −0.5 for ADL and +0.5 for CDL, surprisal, and the interaction between group and surprisal. Surprisal measures were obtained by passing the speech transcripts from our test sets by utterance through the language models (Ecolang_Child test data through the CDL-trained language model, and Ecolang_Adult test data through the ADL-trained language model) and extracting the surprisal for each word in the utterance (Figure 1B and D).

Baseline word duration was included as a control variable to control for the length of the word. Baseline word duration was obtained from the MARY text-to-speech system (MaryTTS; Schröder et al., 2008), following Demberg et al., 2012; Seyfarth, 2014). Utterances from our test sets were sent to the MaryTTS system, and subsequently, word durations were extracted from the realized acoustic parameters for each utterance. Importantly, by realizing entire utterances as opposed to single words, the utterance context is taken into account by the system during speech synthesis. As additional control variables, we included word log frequency (standardized, from SUBTLEX-UK; van Heuven et al., 2014), AoA (standardized; Kuperman et al., 2012), and word position in the utterance. These are all factors that have been shown to affect word durations. Word position was contrast coded using lizContrasts4 functions (Wonnacott et al., 2017), with three levels, coded to medial versus initial, medial versus final, and medial versus isolation, with medial as the reference level (values for the contrasts provided in the online supplemental materials). In addition, we controlled for the effects of the presence or absence of the objects, accounting for the possibility that in the absent condition speakers might speak slower in general. This categorical variable (session) was coded −0.5 for absent condition, and +0.5 for present condition. We also included a variable (order) encoding the order of sessions (i.e., whether the object was present or absent in the first session, coded as 1 when the object was present in the first) section and its interaction with session so that the regression model also takes into account whether data comes from the first or second session. Although the session is not relevant to our purposes here, it has been shown to also affect word duration in CDL (Shi et al., 2022).

For all these control variables, the interaction with group was also included to account for variations in the effects due to differences between CDL and ADL. More specifically, we included an interaction between session, order, and group, accounting for the possibility that speakers could speak slower when object absent condition was first, and they may do so more when addressing young children as opposed to addressing adults. Moreover, we included interaction between word frequency and group, as research suggests that low-frequency words in CDL are generally spoken more slowly compared to the same words in ADL (Tippenhauer et al., 2020). We used the MaryTTS system for establishing baseline word durations. However, it is important to note that this system typically generates word durations characteristic of ADL. Given that CDL often features a slower rate of speech, and extended word durations, as highlighted in research (Fernald & Simon, 1984; Fernald et al., 1989; Soderstrom, 2007), we factored in an interaction between our baseline word duration and group. This adjustment aims to more accurately reflect the distinctive prosodic features of CDL in our analysis. Additionally, we factored in an interaction between AoA and group, aligning with research that

[3] Reliability coding is not carried out in the more traditional way in which coders first agree on criteria and work together through initial disagreements, but in a fully independent manner based only on written coding instructions which explain why while the reliability scores are acceptable, they are lower than often reported.

demonstrates a correlation between the modification in CDL (e.g., word duration) and AoA (Vosoughi et al., 2010). Finally, to further refine our analysis, we included an interaction between word position and group. This decision was supported by studies indicating that word position significantly influences prosodic modulations in CDL (Shi et al., 2022).

We included by-subject random intercepts to account for individual differences in speaking rate, by-subject random slopes for surprisal measures to account for individual differences in sensitivity to surprisal, and by-topic random intercepts and slopes of surprisal to account for differences due to the particular object being talked about. All the numerical predictors were standardized after pooling the CDL and ADL data sets.

Iconic Gestures. We fitted a logistic mixed-effects regression model on pooled child-directed and adult-directed multimodal (gesture and lexical affiliates) data extracted from the ECOLANG corpus. The dependent variable was the presence or absence of an iconic gesture for a gesturable word. We included the same fixed effects and control variables as for the duration analysis except for baseline duration, which was not included. Instead, we entered actual word duration (log-transformed and standardized) as an additional control variable. We included an interaction between session, order, and group. This accounts for the fact that speakers use more gestures in situations where the object is absent (Motamedi et al., 2024), and may do so particularly in the first session. Additionally, this interaction considers the likelihood of increased gesturing when addressing young children compared to adults. We also incorporated an interaction of group and word position to account for the potential that words in different positions in the sentence may be more likely to be accompanied by a gesture, particularly in CDL. Similarly, it has been shown that speakers use more gestures for less common words (Beattie & Shovelton, 2000), and this tendency may be more pronounced in CDL. Additionally, we included an interaction between AoA and group, considering the possibility that speakers might modulate their gesture production differently for words acquired later in both CDL and ADL. Finally, we also factored in the actual word duration and its interaction with group, to control for the possibility that speakers may produce more gestures for words that have longer durations, which again may vary between CDL and ADL.

We included by-subject random intercepts to account for individual differences in the probability of producing an iconic gesture and by-subject random slopes for surprisal measures to account for individual differences in sensitivity to surprisal. In addition, we included by-topic random intercepts to account for differences in gesturability of the object being talked about, and by-topic random slopes of surprisal because participants' sensitivity to surprisal may differ depending on the object.

Analysis 2: Are Adjustments (In)Dependent Across Modalities?

We fitted a linear mixed-effects regression model on the multimodal (gesture and gesturable words only) pooled CDL and ADL data. In this model, the dependent variable was observed log-transformed word duration in milliseconds, and the experimental variables were surprisal (standardized), gesture presence (centered), and their interaction. In addition, the model included the following control variables: log-transformed and standardized baseline word duration, word log frequency (standardized), AoA (standardized), word position in the utterance, session, and session order, coded as in Analysis 1. Furthermore, we included an interaction between session, order, and group, as well as interactions between word position and group, AoA and group, word frequency and group, and word duration and group.

We included by-subject random intercepts to account for individual differences in gesture production, by-subject random slopes for surprisal measures to account for individual differences in sensitivity to surprisal, and by-topic random intercepts and slopes of surprisal to account for differences due to the particular object being talked about. All the numerical predictors were standardized after pooling the CDL and ADL data sets.

Transparency and Openness

The number of participants was determined opportunistically as the number of participants in the corpus for whom annotation of the speech and gesture was completed at the time of the study. All data files, analysis scripts, and other materials pertaining to the study can be found at the project's OSF page https://osf.io/ugs28/. The study was not preregistered.

Results

Analysis 1: Does Surprisal Predict Multimodal Adjustments in CDL and ADL?

Word Durations

Table 1 summarizes the key results from the analysis of word durations on pooled data from CDL and ADL. Full analysis results are available in the online supplemental materials (data and code also available at the project's OSF page https://osf.io/ugs28/). We found overall longer durations in CDL than in ADL, as expected based on the previous literature on CDL. Crucially, after controlling for baseline word duration, AoA, word position, word frequency, presence of the object, and session order, there is a significant positive effect of surprisal on word durations. As Figure 2 (right) illustrates, more surprising words are spoken more slowly than words that are less surprising, both in CDL and ADL, although the effect is slightly weaker in CDL. These results were confirmed by likelihood-ratio tests that revealed a better fit to data of a regression model with than without the Surprisal × Group interaction, $\chi^2(1) = 6.40$, $p < .02$, as well as a better fit of a model with than without the surprisal fixed effect, $\chi^2(1) = 139.9$, $p < .00001$.

Iconic Gestures

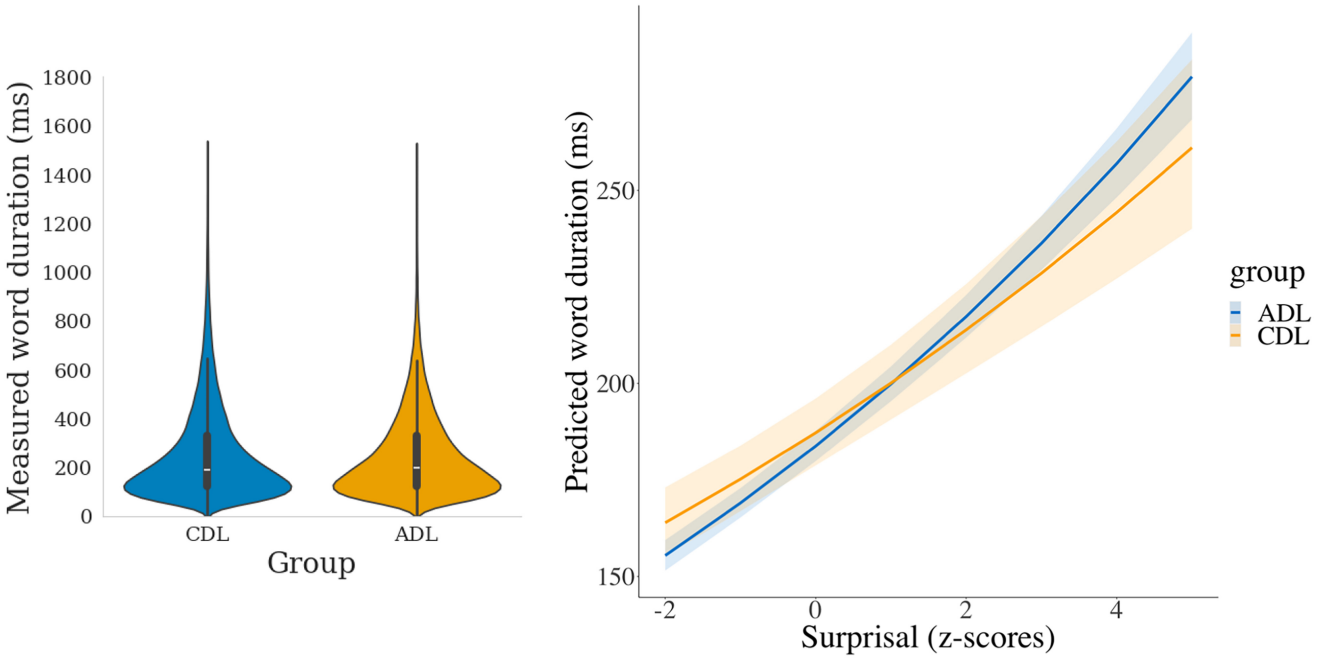
Table 2 summarizes the results of the logistic regression analysis of iconic gestures on pooled data from child-directed and adult-directed language. Full analysis results are available in the online supplemental materials (full data and code for the analysis can also be found on the

Table 1
Summary of Regression Model Fitted to Word Duration

Word duration	Estimate	Estimated error	*t*
Intercept	5.460	0.015	353.83
Group (CDL vs. ADL)	0.022	0.030	0.73
Surprisal	0.075	0.003	21.52
Surprisal × Group	−0.017	0.007	−2.57

Note. CDL = child-directed language; ADL = adult-directed language.

Figure 2
Word Duration



Note. Left panel: Distribution of word durations (ms) in the corpus (the white dot indicates the median, the central thick bar denotes the interquartile range, and the thin line represents 1.5 times the interquartile range). The plot's width at various points reflects the density distribution of data. Right panel: Predicted word durations (ms) as a function of surprisal in ADL and CDL. Shaded areas indicate standard errors. CDL = child-directed language; ADL = adult-directed language. See the online article for the color version of this figure.

project's OSF page https://osf.io/ugs28/). Overall, there is a main effect of Group in that gesturable words (i.e., those that appear in the corpus at least once as lexical affiliate) are more likely to be accompanied by a gesture in ADL than in CDL (Figure 3, left). After taking into account all the control variables, there was a reliable effect of surprisal on gesture probability: words with higher surprisal have a higher probability of being accompanied by a gesture. In contrast to word duration, there is no significant interaction between surprisal and group (Figure 3, right). These results were confirmed by likelihood-ratio tests that revealed no significant difference in fit to data of regression model with and without the Surprisal × Group interaction, $\chi^2(1) = 0.27$, $p = .60$, but better fit of a model with than without the surprisal fixed effect, $\chi^2(1) = 49.9$, $p < .00001$.

Analysis 2: Are Adjustments (In)Dependent Across Modalities?

Table 3 summarizes the core results from the analysis that predicted word durations based on surprisal and its interaction with gesture presence. Full analysis results are available in the online supplemental materials (full data and code for the analysis can also be found on the project's OSF page https://osf.io/ugs28/). The analysis used pooled data from both CDL and ADL. A crucial observation is that the surprisal effect is reduced when a gesture accompanies speech, likelihood-ratio test, $\chi^2(1) = 14.5$, $p < .0002$. This finding suggests that speakers may favor modulating one mode of communication—either gesture or word duration—rather than simultaneously modulating both.

However, surprisal appears less impacted by concurrent gesturing in the CDL rather than in the ADL, likelihood-ratio test, $\chi^2(1) = 5.90$, $p < .02$, data set. Notably, a separate analysis of the ADL and CDL data revealed that the interaction between surprisal and gesture was not statistically significant for CDL (estimate = −0.012; $SE = 0.013$; $t = -0.90$). Thus, for CDL, our data suggest that speakers operate the two communication channels independently. Note, however, that lack of statistical significance does not provide concrete evidence for independence; it simply implies an absence of compelling evidence for a dependency relationship (Figure 4).

Table 2
Summary of Regression Model Fitted to Gesture Presence

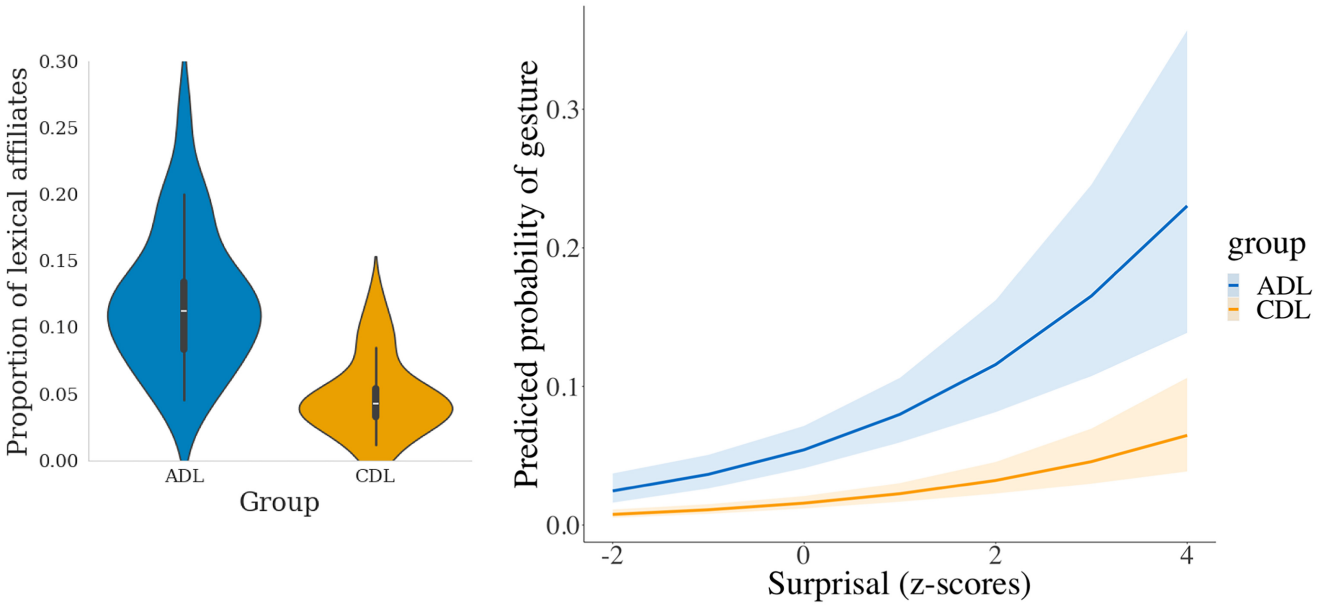
Gesture presence	Estimate	Estimated error	z	p
Intercept	−3.866	0.155	−24.94	<.00001
Group (CDL vs. ADL)	−1.149	0.285	−4.03	<.0001
Surprisal	0.389	0.051	7.69	<.00001
Surprisal × Group	−0.047	0.089	−0.53	.60

Note. CDL = child-directed language; ADL = adult-directed language.

Discussion

In face-to-face communication, speakers can use multimodal cues to communicate efficiently, that is, with minimal effort and in a successful manner. For these multimodal cues to subserve communicative efficiency, we should find that speakers adapt them on the fly to the predictability of words in context. We quantified predictability in terms of surprisal obtained from language models trained on corpora of CDL and ADL. We then tested whether surprisal predicts word

Figure 3
Iconic Gestures



Note. Left panel: Proportions of words that appear at least once as lexical affiliate of an iconic gesture (referred as gesturable words in the text) per participant for ADL and CDL (the white dot indicates the median, the central thick bar denotes the interquartile range, and the thin line represents 1.5 times the interquartile range). The plot's width at various points reflects the density distribution of data. Right panel: Predicted probability of gesture as a function of surprisal in child-directed and adult-directed language. Shaded areas indicate standard errors. ADL = adult-directed language; CDL = child-directed language. See the online article for the color version of this figure.

durations and the presence of gestures using multimodal data taken from the new ECOLANG corpus of dyadic interaction.

We found that, overall, surprisal predicted whether speakers make words longer and produce iconic gestures, for both CDL and ADL. We further found some differences depending upon the addressee with a slightly weaker effect of surprisal on duration in CDL than ADL. Moreover, the surprisal effect on duration was more strongly reduced in the presence of a gesture in ADL than in CDL; in fact, this reduction was only statistically significant in ADL. Below we discuss these results in turn.

Efficiency in Multimodal Communication

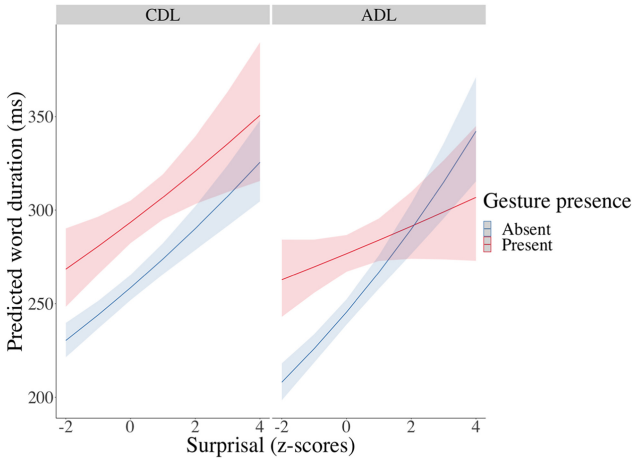
It has been proposed that human language is optimized for communicative efficiency (Gibson et al., 2019; Levshina, 2022). Some previous work has shown this to be the case for a number of structural properties of language such as lexical realization of concepts (Gibson et al., 2017; Kemp & Regier, 2012), word order (Gibson et al., 2013), and syntactic dependencies (Futrell et al., 2015) among others.

Table 3
Summary of Regression Model Fitted to Word Duration Data, With Gesture Presence Included as a Predictor

Word duration	Estimate	Estimated error	*t*
Intercept	5.647	0.0144	393.41
Group (CDL vs. ADL)	0.130	0.028	4.66
Surprisal	0.070	0.007	10.71
Gesture presence	0.123	0.010	12.65
Surprisal × Group	−0.025	0.012	−2.14
Gesture presence × Group	0.007	0.194	0.37
Surprisal × Gesture Presence	−0.035	0.009	−3.87
Surprisal × Group × Gesture presence	0.044	0.018	2.43

Note. CDL = child-directed language; ADL = adult-directed language.

Figure 4
Predicted Word Duration as a Function of Surprisal and Gesture Presence in Child-Directed (Left Panel) and Adult-Directed (Right Panel) Language



Note. Shaded areas indicate standard errors. CDL = child-directed language; ADL = adult-directed language. See the online article for the color version of this figure.

This work looks at language as a population-level system focusing on structural properties. Efficiency in usage, however, should be manifested first and foremost in situated language, namely in the moment of communication (Murgiano et al., 2021). In situated language, the units on which the pressure for efficiency operates go beyond speech including a range of multimodal behaviors that communicative partners engage in when talking (Holler & Levinson, 2019). Some previous work has investigated adjustments of word duration, in adult-directed speech, as an example of how efficiency exerts pressure on communication in the moment (Demberg et al., 2012; Seyfarth, 2014). These studies found that speakers' modulations of word duration in ADL fulfill the requirements for efficiency. Namely, speakers tend to shorten words that are highly predictable (decreasing effort) and lengthen words that are less predictable (increasing likelihood of communicative success). It has further been suggested that our processing system operates in such a way to distribute information in a uniform manner over time (the uniform information density hypothesis, UID) to reach communicative success while at the same time avoiding excessive cognitive load (Aylett & Turk, 2006; Levy & Jaeger, 2007; Pluymaekers et al., 2005). UID nicely captures phenomena such as modulations of word duration according to predictability. Here we replicated these findings for ADL and we extended them to CDL.

However, word duration is only one aspect (temporal adjustments of the vocal modality) of the signal that speakers can adjust. Gestural behaviors can provide an additional signal (at least partially redundant with the speech) over a different modality (a manual channel that operates in the spatial as well as temporal dimensions) that speakers can exploit, at the service of communicative efficiency. One previous study has provided initial evidence that gesture production, in adult-directed language, is modulated by the predictability of the corresponding word on a very limited set of words and a small number of subjects (Beattie & Shovelton, 2000). No previous study has addressed this question in child-directed language. We find that surprisal modulates the probability of gesture production both in CDL and ADL, as it did for word durations.

Crucially, if efficiency is distributed across different channels, speakers may choose how to distribute effort. They may, for example, increase duration and produce a gesture for a more surprising word to maximize communication success at the expense of increasing cognitive/production load. Or, they may, for example, increase word duration but refrain from producing a gesture, such that whereas communicative success is increased, the load is only increased in one but not the other channel. Alternatively, surprisal may modulate the two channels independently. No previous study has asked whether and how adjustments across modalities are independent. We find a general tendency for speakers talking to other adults to avoid modulating both channels, such that if they produce a gesture, they are less likely to also extend the word's duration. This effect, however, is clear in ADL only as for CDL, we do not find evidence for dependency between the two channels.

Bringing together adjustments in word duration and gesture probability, our study is the first to investigate communicative efficiency in multimodal language. Our results are compatible with the idea that speakers adjust their multimodal behaviors while speaking, as predicted by communicative efficiency, across the vocal and gestural modalities. This is not surprising as the ecology of communication is face to face in which both the vocal and gestural modalities are available. Above, we introduced the UID hypothesis as a mechanism for implementing communicative efficiency. While this hypothesis has been framed in terms of distribution over time, our results suggest that information can be distributed across different channels both in the temporal as well as in the spatial dimension.

How Audience Characteristics Affect Efficiency

In line with the predictions from communicative efficiency, we found clear overall effects of surprisal both on CDL as well as ADL. For word durations, we found that effects were modulated by the characteristics of the addressee. First, we found a somewhat greater effect of surprisal on word duration in ADL than in CDL. More specifically, in Analysis 2 (which is limited to gesturable words), we see that CDL tends to use longer durations across the board. As already discussed in the introductory part, a slower speaking rate is a prime characteristic of child-directed speech that decreases around the age of 2, but persists well into the age range considered here (with substantial variability, e.g., Ko, 2012). Along with other properties of child-directed speech (or "parentese"), it has been argued to support different aspects of language development and processing not directly related to predictability (e.g., Golinkoff et al., 2015; Ma et al., 2011; Thiessen et al., 2005). A previous study that addressed word duration modulations as a function of predictability comparing ADL and CDL obtained different results (Pate & Goldwater, 2015). That study found that word frequency, predictability from the preceding context, and predictability from the following context predict word durations in ADL; however, only word frequency and predictability from the following context predicted word durations in CDL. Here we found an effect of predictability from the preceding context in both. There are important differences between the two studies. First, Pate and Goldwater investigated infant-direct language whereas we focus on the language directed to older (3–4 years) children. CDL to children of this age still maintains some of the properties of the register used with infants; however, it is far richer in terms of vocabulary and types of utterances. Second, we modeled predictability on the basis of CDL and ADL corpora, rather than a single corpus. Therefore, our measures more closely match predictability from the perspective of the comprehender. Finally, our ECOLANG data elicitation procedure is closely matched between ADL and CDL, making the data sets comparable whereas this is not the case in Pate and Goldwater's work.

While there is abundant literature on prosodic modulation both in CDL and ADL and their roles in language development and processing, more limited work exists concerning iconic gesture production. Recent work on the ECOLANG corpus (Donnellan et al., 2022), and from other groups (Ter Bekke et al., 2020), found that speakers tend to produce iconic gestures just before (approximately 300 ms) the corresponding lexical affiliate. The same timing relationship is found in CDL for words that the child knows (Donnellan et al., 2022; Wang et al., 2023). This timing relationship allows the comprehender to use the gesture to predict the upcoming word and therefore is compatible with the hypothesis that iconic gestures can support communicative success across ADL and CDL. Note that the main effect of group here cannot be taken to show a larger probability to gesture in ADL than in CDL, because the analysis was carried out only on a subset of words (the gesturable word tokens, namely those words accompanied at least once by a gesture).

Crucially, we found that speakers were less likely to combine modulations of duration and production of gestures when talking to an adult than when talking to a child. Thus, when talking to another adult, speakers were "most efficient" in their use of

multimodal language, increasing effort to the service of communicative success primarily in one modality. This does not mean that modulations do not occur at all in the other modality, but that they tend to be more focused on one modality only. However, when talking to a child, this was not the case: modulations of word durations and gesture production were independent. This result, again, may be accounted for by the fact that modulations of word duration may serve different functions in CDL, as described above.

When considering audience effects in our study, it is important to take into account that our surprisal estimates were obtained from a language model trained on either CDL or ADL, depending on whether we analyzed language directed toward a child or an adult. This implements the assumption that speakers will adjust their internal language model to their addressee, that is, speakers' predictions about what is unexpected to their interlocutor depend on their belief about the listener's language knowledge. However, it is also possible—although less likely—that speakers do not adapt their modulations of speech rate and gesture use in this manner, that is, they use one and the same internal language model to estimate how unexpected a word is to the listener. To investigate how robust our results are to the assumption that speakers adapt their language model to the listener's age, we reran our analyses using surprisal estimates that combine the two language models by averaging the probabilities from the ADL- and CDL-trained neural networks. These analyses, too, revealed surprisal effects on duration, regression coefficient, $b = .085$, $t = 23.54$; likelihood-ratio test, $\chi^2(1) = 161.0$, $p < .00001$, and gesture probability, regression coefficient, $b = .420$, $z = 8.20$, $p < .00001$; likelihood-ratio test, $\chi^2(1) = 56.6$, $p < .00001$. However, there was no longer a significant Surprisal × Group interaction on duration, regression coefficient, $b = -.007$, $t = -0.97$; likelihood-ratio test, $\chi^2(1) = 0.93$, $p > .3$, nor on gesture probability, regression coefficient, $b = .162$, $z = 1.82$, $p = .07$; likelihood-ratio test, $\chi^2(1) = 3.20$, $p = .07$. Analysis 2 also did not reveal the three-way interaction between surprisal, group, and gesture presence on duration, regression coefficient, $b = .026$, $t = 1.45$; likelihood-ratio test, $\chi^2(1) = 2.10$, $p = .15$. In short, all differences in surprisal effect between ADL and CDL that we found, depend on our assumption that speakers take an age-appropriate perspective when predicting how unexpected a word will be to the listener.

Are Multimodal Adaptations Serving Language Production, Comprehension, or Communication?

As described in the introductory part, adjustments of word duration and production of gestures have been shown to correlate with variables such as lexical frequency and AoA, which are known to affect word recognition (Dahan & Magnuson, 2006), but also importantly difficulty of word retrieval and production (Arnold, 2008; Aylett & Turk, 2004; Florian Jaeger, 2010; Gahl et al., 2012; Watson et al., 2008). Speakers could simply make these adjustments to support their production-internal processes with no consideration of the comprehender. Indeed, we found that word frequency and AoA significantly predict the multimodal adjustments in our corpus (see the online supplemental materials). However, this is clearly not the whole story as we find effects of surprisal above and beyond these lexical variables. Thus, multimodal adjustments can serve communication supporting both the speaker and the comprehender. More speculatively, one may argue that these adjustments have been maintained in the evolution of language because they play a role in supporting both communication partners.

Efficiency as defined within an information-theoretic approach applies to a communicative system that includes both the speaker and the comprehender. Effort management can apply to both production and comprehension (i.e., shorter words require less effort to produce and to comprehend, a gesture takes effort to produce and to comprehend); communicative success is also to the benfit of both producer and comprehender. It remains an open question whether in this framework a definition of predictability that takes into account utterances produced by both partners (rather than on the utterances produced by one speaker only) may prove to be an even better predictor for the different multimodal behaviors that characterize our communication. In general, this framework encourages an approach in which we move beyond considering the speaker and the comprehender as separate entities that are investigated independently (and language production and comprehension as two separate subfields).

Conclusions

Communicative efficiency has been shown to account for key structural properties of language. Here we have argued that, as human communication is situated and face to face, efficiency should be defined over the ensemble of vocal and gestural behaviors that ubiquitously accompany speech, rather than just linguistic units. We have found that communicative efficiency can account for the production of multimodal adjustments, regardless of whether the comprehender is an adult or a young child and therefore whether the comprehender is only processing or also still learning language. This framework provides a parsimonious explanation of different communicative phenomena (linguistic and nonlinguistic) that are traditionally investigated separately.

Constraints on Generality

The current study uses data from the ECOLANG corpus of dyadic communication which includes both U.K. and U.S. English speakers (Gu et al., n.d.). The results obtained here are most directly generalizable to child and adult English speakers. Note, however, that because related findings have been previously reported for prosodic modulation in other languages (such as German and Swedish) and iconic gestures have been documented across a vast array of the world's languages, we would expect the present results to generalize to other Indo-European languages.

References

Acredolo, L. P., & Goodwyn, S. (1988). Symbolic gesturing in normal infants. *Child Development*, *59*(2), 450–466. https://doi.org/10.2307/1130324

Alibali, M. W., Heath, D. C., & Myers, H. J. (2001). Effects of visibility between speaker and listener on gesture production: Some gestures are meant to be seen. *Journal of Memory and Language*, *44*(2), 169–188. https://doi.org/10.1006/jmla.2000.2752

Arnold, J. E. (2008). Reference production: Production-internal and addressee-oriented processes. *Language and Cognitive Processes*, *23*(4), 495–527. https://doi.org/10.1080/01690960801920099

Aurnhammer, C., & Frank, S. L. (2019). Evaluating information-theoretic measures of word prediction in naturalistic sentence reading. *Neuropsychologia*, *134*, Article 107198. https://doi.org/10.1016/j.neuropsychologia.2019.107198

Aussems, S., & Kita, S. (2019). Seeing iconic gestures while encoding events facilitates children's memory of these events. *Child Development*, *90*(4), 1123–1137. https://doi.org/10.1111/cdev.12988

Aylett, M., & Turk, A. (2004). The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech*, *47*(1), 31–56. https://doi.org/10.1177/00238309040470010201

Aylett, M., & Turk, A. (2006). Language redundancy predicts syllabic duration and the spectral characteristics of vocalic syllable nuclei. *The Journal of the Acoustical Society of America*, *119*(5 Pt 1), 3048–3058. https://doi.org/10.1121/1.2188331

Barry, C., Hirsh, K., Johnston, R., & Williams, C. L. (2001). *Age of acquisition, word frequency, and the locus of repetition priming of picture naming*. https://doi.org/10.1006/JMLA.2000.2743

Beattie, G., & Shovelton, H. (2000). Iconic hand gestures and the predictability of words in context in spontaneous speech. *British Journal of Psychology*, *91*(4), 473–491. https://doi.org/10.1348/000712600161943

Bell, A. (1984). Language style as audience design. *Language in Society*, *13*(2), 145–204. https://doi.org/10.1017/S004740450001037X

Ben-Aderet, T., Gallego-Abenza, M., Reby, D., & Mathevon, N. (2017). Dog-directed speech: Why do we use it and do dogs pay attention to it? *Proceedings. Biological Sciences*, *284*(1846), Article 20162429. https://doi.org/10.1098/rspb.2016.2429

Boersma, P., & Weenink, D. (1992–2022). *Praat: Doing phonetics by computer* (Version 6.3.17) [Computer software]. https://www.praat.org

Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(6), 1482–1493. https://doi.org/10.1037/0278-7393.22.6.1482

Cho, K., van Merrienboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). *Learning phrase representations using RNN encoder–decoder for statistical machine translation*. arXiv:1406.1078 [Cs, Stat]. https://arxiv.org/abs/1406.1078

Clark, H. H., & Fox Tree, J. E. (2002). Using uh and um in spontaneous speaking. *Cognition*, *84*(1), 73–111. https://doi.org/10.1016/S0010-0277(02)00017-3

Cooper, R. P., & Aslin, R. N. (1990). Preference for infant-directed speech in the first month after birth. *Child Development*, *61*(5), 1584–1595. https://doi.org/10.2307/1130766

Cristia, A. (2013). Input to language: The phonetics and perception of infant-directed speech. *Language and Linguistics Compass*, *7*(3), 157–170. https://doi.org/10.1111/lnc3.12015

Cutler, A. (1996). Prosody and the word boundary problem. In J. L. Morgan & K. Demuth (Eds.), *Signal to syntax: Bootstrapping from speech to grammar in early acquisition* (pp. 87–99). Lawrence Erlbaum Associates.

Dahan, D., & Magnuson, J. S. (2006). Spoken-word recognition. In M. J. Traxler & M. A. Gernsbacher (Eds.), *Handbook of psycholinguistics* (pp. 249–283). Academic Press.

Demberg, V., Sayeed, A., Gorinski, P., & Engonopoulos, N. (2012). Syntactic surprisal affects spoken word duration in conversational contexts. In J. Tsujii, J. Henderson, & M. Paşca (Eds.), *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning* (pp. 356–367). Association for Computational Linguistics.

De Ruiter, J. P. (2000). The production of gesture and speech. In D. McNeill (Ed.), *Language and gesture* (pp. 248–311). Cambridge University Press.

Donnellan, E., Ozder, L. E., Man, H., Grzyb, B., Gu, Y., & Vigliocco, G. (2022). Timing relationships between representational gestures and speech: A corpus based investigation. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *44*(44), 2052–2058. https://escholarship.org/uc/item/7w349725

Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, *14*(2), 179–211. https://doi.org/10.1207/s15516709cog1402_1

Estes, K. G., & Hurley, K. (2013). Infant-directed prosody helps infants map sounds to meanings. *Infancy: The Official Journal of the International Society on Infant Studies*, *18*(5). https://doi.org/10.1111/infa.12006

Fernald, A. (1992). Meaningful melodies in mothers' speech to infants. In H. Papoušek, U. Jürgens, & M. Papoušek (Eds.), *Nonverbal vocal communication: Comparative and developmental approaches* (pp. 262–282). Editions de la Maison des Sciences de l'Homme; Cambridge University Press.

Fernald, A. (2000). Speech to infants as hyperspeech: Knowledge-driven processes in early word recognition. *Phonetica*, *57*(2–4), 242–254. https://doi.org/10.1159/000028477

Fernald, A., & Simon, T. (1984). Expanded intonation contours in mothers' speech to newborns. *Developmental Psychology*, *20*(1), 104–113. https://doi.org/10.1037/0012-1649.20.1.104

Fernald, A., Taeschner, T., Dunn, J., Papousek, M., de Boysson-Bardies, B., & Fukui, I. (1989). A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants. *Journal of Child Language*, *16*(3), 477–501. https://doi.org/10.1017/S0305000900010679

Flevares, L. M., & Perry, M. (2001). How many do you see? The use of nonspoken representations in first-grade mathematics lessons. *Journal of Educational Psychology*, *93*(2), 330–345. https://doi.org/10.1037/0022-0663.93.2.330

Florian Jaeger, T. (2010). Redundancy and reduction: Speakers manage syntactic information density. *Cognitive Psychology*, *61*(1), 23–62. https://doi.org/10.1016/j.cogpsych.2010.02.002

Frank, S. L., Otten, L. J., Galli, G., & Vigliocco, G. (2015). The ERP response to the amount of information conveyed by words in sentences. *Brain and Language*, *140*, 1–11. https://doi.org/10.1016/j.bandl.2014.10.006

Frank, S. L., Trompenaars, T., & Vasishth, S. (2016). Cross-linguistic differences in processing double-embedded relative clauses: Working-memory constraints or language statistics? *Cognitive Science*, *40*(3), 554–578. https://doi.org/10.1111/cogs.12247

Friedman, L. A. (1977). *On the other hand: New perspectives on American sign language*. Academic Press.

Futrell, R., Mahowald, K., & Gibson, E. (2015). Large-scale evidence of dependency length minimization in 37 languages. *Proceedings of the National Academy of Sciences*, *112*(33), 10336–10341. https://doi.org/10.1073/pnas.1502134112

Gahl, S., Yao, Y., & Johnson, K. (2012). Why reduce? Phonological neighborhood density and phonetic reduction in spontaneous speech. *Journal of Memory and Language*, *66*(4), 789–806. https://doi.org/10.1016/j.jml.2011.11.006

Galati, A., & Brennan, S. E. (2014). Speakers adapt gestures to addressees' knowledge: Implications for models of co-speech gesture. *Language, Cognition and Neuroscience*, *29*(4), 435–451. https://doi.org/10.1080/01690965.2013.796397

Gerhand, S., & Barry, C. (1999). Age-of-acquisition and frequency effects in speeded word naming. *Cognition*, *73*(2), B27–B36. https://doi.org/10.1016/S0010-0277(99)00052-9

Gibson, E., Futrell, R., Jara-Ettinger, J., Mahowald, K., Bergen, L., Ratnasingam, S., Gibson, M., Piantadosi, S. T., & Conway, B. R. (2017). Color naming across languages reflects color use. *Proceedings of the National Academy of Sciences*, *114*(40), 10785–10790. https://doi.org/10.1073/pnas.1619666114

Gibson, E., Futrell, R., Piantadosi, S. P., Dautriche, I., Mahowald, K., Bergen, L., & Levy, R. (2019). How efficiency shapes human language. *Trends in Cognitive Sciences*, *23*(5), 389–407. https://doi.org/10.1016/j.tics.2019.02.003

Gibson, E., Piantadosi, S. T., Brink, K., Bergen, L., Lim, E., & Saxe, R. (2013). A noisy-channel account of crosslinguistic word-order variation. *Psychological Science*, *24*(7), 1079–1088. https://doi.org/10.1177/0956797612463705

Goldin-Meadow, S. (1999). The role of gesture in communication and thinking. *Trends in Cognitive Sciences*, *3*(11), 419–429. https://doi.org/10.1016/S1364-6613(99)01397-2

Golinkoff, R. M., Can, D. D., Soderstrom, M., & Hirsh-Pasek, K. (2015). (Baby) talk to me: The social context of infant-directed speech and its effects on early language acquisition. *Current Directions in Psychological Science*, *24*(5), 339–344. https://doi.org/10.1177/0963721415595345

Goodkind, A., & Bicknell, K. (2018). Predictive power of word surprisal for reading times is a linear function of language model quality. In A. Sayeed, C. Jacobs, T. Linzen, & M. van Schijndel (Eds.), *Proceedings of the 8th Workshop on Cognitive Modeling and Computational Linguistics (CMCL 2018)* (pp. 10–18). Association for Computational Linguistics. https://doi.org/10.18653/v1/W18-0102

Goodrich, W., & Hudson Kam, C. L. (2009). Co-speech gesture as input in verb learning. *Developmental Science*, *12*(1), 81–87. https://doi.org/10.1111/j.1467-7687.2008.00735.x

Grice, H. P. (1989). *Studies in the way of words*. Harvard University Press.

Gu, Y., Donnellan, E., Grzyb, B., Brekelmans, G., Murgiano, M., Brieke, R., Perniss, P., & Vigliocco, G. (n.d.). *The ECOLANG multimodal corpus of adult–child and adult–adult conversation*.

Gulordava, K., Bojanowski, P., Grave, E., Linzen, T., & Baroni, M. (2019). Colorless green recurrent networks dream hierarchically. *Proceedings of the Society for Computation in Linguistics*, 2, Article 48. https://scholarworks.umass.edu/scil/vol2/iss1/48

Hadar, U., & Butterworth, B. (2009). Iconic gestures, imagery, and word retrieval in speech. *Semiotica*, *115*(1–2), 147–172. https://doi.org/10.1515/semi.1997.115.1-2.147

Hale, J. (2001, June 1–7). A probabilistic early parser as a psycholinguistic model. In *Proceedings of the second meeting of the north American chapter of the association for computational linguistics on language technologies, Pittsburgh, Pennsylvania*. Association for Computational Linguistics.

Hall, K. C., Hume, E., Jaeger, T. F., & Wedel, A. (2018). The role of predictability in shaping phonological patterns. *Linguistics Vanguard*, *4*(S2), Article 20170027. https://doi.org/10.1515/lingvan-2017-0027

Haspelmath, M. (2021). Towards standardization of morphosyntactic terminology for general linguistics. In L. Alfieri, G. F. Arcodia, & P. Ramat (Eds.), *Linguistic categories, language description and linguistic typology* (pp. 35–58). John Benjamins Publishing. https://doi.org/10.1075/tsl.132.02has

Hintz, F., Khoe, Y. H., Strauß, A., Psomakas, A. J. A., & Holler, J. (2023). Electrophysiological evidence for the enhancement of gesture-speech integration by linguistic predictability during multimodal discourse comprehension. *Cognitive, Affective, & Behavioral Neuroscience*, *23*(2), 340–353. https://doi.org/10.3758/s13415-023-01074-8

Hoetjes, M., Koolen, R., Goudbeek, M., Krahmer, E., & Swerts, M. (2015). Reduction in gesture during the production of repeated references. *Journal of Memory and Language*, *79–80*, 1–17. https://doi.org/10.1016/j.jml.2014.10.004

Holle, H., & Gunter, T. C. (2007). The role of iconic gestures in speech disambiguation: ERP evidence. *Journal of Cognitive Neuroscience*, *19*(7), 1175–1192. https://doi.org/10.1162/jocn.2007.19.7.1175

Holler, J., & Levinson, S. C. (2019). Multimodal language processing in human communication. *Trends in Cognitive Sciences*, *23*(8), 639–652. https://doi.org/10.1016/j.tics.2019.05.006

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

Humphreys, G., Riddoch, M., & Quinlan, P. (1988). *Cascade processes in picture identification*. https://doi.org/10.1080/02643298808252927

Iverson, J. M., Capirci, O., Longobardi, E., & Caselli, M. C. (1999). Gesturing in mother–child interactions. *Cognitive Development*, *14*(1), 57–75. https://doi.org/10.1016/S0885-2014(99)80018-5

Jacobs, N., & Garnham, A. (2007). The role of conversational hand gestures in a narrative task. *Journal of Memory and Language*, *56*(2), 291–303. https://doi.org/10.1016/j.jml.2006.07.011

Jaeger, T. F., & Tily, H. (2011). On language 'utility': Processing complexity and communicative efficiency. *Wiley Interdisciplinary Reviews. Cognitive Science*, *2*(3), 323–335. https://doi.org/10.1002/wcs.126

Jescheniak, J. D., & Levelt, W. J. M. (1994). Word frequency effects in speech production: Retrieval of syntactic information and of phonological form. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(4), 824–843. https://doi.org/10.1037/0278-7393.20.4.824

Kelly, S., Byrne, K., & Holler, J. (2011). Raising the ante of communication: Evidence for enhanced gesture use in high stakes situations. *Information*, *2*(4), 579–593. https://doi.org/10.3390/info2040579

Kemp, C., & Regier, T. (2012). Kinship categories across languages reflect general communicative principles. *Science*, *336*(6084), 1049–1054. https://doi.org/10.1126/science.1218811

Kendon, A. (2004). *Gesture: Visible action as utterance*. Cambridge University Press. https://doi.org/10.1017/CBO9780511807572

Kita, S., Alibali, M. W., & Chu, M. (2017). How do gestures influence thinking and speaking? The gesture-for-conceptualization hypothesis. *Psychological Review*, *124*(3), 245–266. https://doi.org/10.1037/rev0000059

Kita, S., & Ozyurek, A. (2003). What does cross-linguistic variation in semantic coordination of speech and gesture reveal? Evidence for an interface representation of spatial thinking and speaking. *Journal of Memory and Language*, *48*(1), 16–32. https://doi.org/10.1016/S0749-596X(02)00505-3

Ko, E.-S. (2012). Nonlinear development of speaking rate in child-directed speech. *Lingua*, *122*(8), 841–857. https://doi.org/10.1016/j.lingua.2012.02.005

Krauss, R., Chen, Y., & Gottesman, R. (2000). Lexical gestures and lexical access: A process model. In D. McNeill (Ed.), *Language and gesture* (pp. 261–283). Cambridge University Press.

Kuhl, P. K., Andruski, J. E., Chistovich, I. A., Chistovich, L. A., Kozhevnikova, E. V., Ryskina, V. L., Stolyarova, E. I., Sundberg, U., & Lacerda, F. (1997). Cross-language analysis of phonetic units in language addressed to infants. *Science*, *277*(5326), 684–686. https://doi.org/10.1126/science.277.5326.684

Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behavior Research Methods*, *44*(4), 978–990. https://doi.org/10.3758/s13428-012-0210-4

Kurumada, C., & Jaeger, T. F. (2015). Communicative efficiency in language production: Optional case-marking in Japanese. *Journal of Memory and Language*, *83*, 152–178. https://doi.org/10.1016/j.jml.2015.03.003

Lausberg, H., & Sloetjes, H. (2009). Coding gestural behavior with the NEUROGES-ELAN system. *Behavior Research Methods*, *41*(3), 841–849. https://doi.org/10.3758/BRM.41.3.841

Levshina, N. (2022). Frequency, informativity and word length: Insights from typologically diverse corpora. *Entropy*, *24*(2), Article 2. https://doi.org/10.3390/e24020280

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, *106*(3), 1126–1177. https://doi.org/10.1016/j.cognition.2007.05.006

Levy, R., & Jaeger, T. F. (2007). Speakers optimize information density through syntactic reduction. In B. Schölkopf, J. Platt, & T. Hoffman (Eds.), *Advances in neural information processing systems (NIPS)* (Vol. 19, pp. 849–856). MIT Press.

Love, R., Dembry, C., Hardie, A., Brezina, V., & McEnery, T. (2017). The spoken BNC2014: Designing and building a spoken corpus of everyday conversations. *International Journal of Corpus Linguistics*, *22*(3), 319–344. https://doi.org/10.1075/ijcl.22.3.02lov

Lowder, M. W., Choi, W., Ferreira, F., & Henderson, J. M. (2018). Lexical predictability during natural reading: Effects of surprisal and entropy reduction. *Cognitive Science*, *42*(S4), 1166–1183. https://doi.org/10.1111/cogs.12597

Ma, W., Golinkoff, R. M., Houston, D. M., & Hirsh-Pasek, K. (2011). Word learning in infant- and adult-directed speech. *Language Learning and Development*, *7*(3), 185–201. https://doi.org/10.1080/15475441.2011.579839

Mahowald, K., Fedorenko, E., Piantadosi, S. T., & Gibson, E. (2013). Info/information theory: Speakers choose shorter words in predictive contexts. *Cognition*, *126*(2), 313–318. https://doi.org/10.1016/j.cognition.2012.09.010

Mani, N., & Huettig, F. (2012). Prediction during language processing is a piece of cake—But only for skilled producers. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(4), 843–847. https://doi.org/10.1037/a0029284

McAuliffe, M., Socolof, M., Mihuc, S., Wagner, M., & Sonderegger, M. (2017, August 20–24). *Montreal forced aligner: Trainable text-speech alignment using kaldi*. INTERSPEECH 2017, Stockholm, Sweden. https://doi.org/10.21437/INTERSPEECH.2017-1386

McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. University of Chicago Press.

Merkx, D., & Frank, S. L. (2021). Human sentence processing: Recurrence or attention? In E. Chersoni, N. Hollenstein, C. Jacobs, Y. Oseki, L. Prévot, & E. Santus (Eds.), *Proceedings of the workshop on cognitive modeling and computational linguistics* (pp. 12–22). Association for Computational Linguistics. https://doi.org/10.18653/v1/2021.cmcl-1.2

Meylan, S. C., & Griffiths, T. L. (2021). The challenges of large-scale, web-based language datasets: Word length and predictability revisited. *Cognitive Science*, *45*(6), Article e12983. https://doi.org/10.1111/cogs.12983

Motamedi, Y., Murgiano, M., Grzyb, B., Gu, Y., Kewenig, V., Brieke, R., Donnellan, E., Marshall, C., Wonnacott, E., Perniss, P., & Vigliocco, G. (2024). *Language development beyond the here-and-now: Iconicity and displacement in child-directed communication. Child Development*. Advance online publication. https://doi.org/10.1111/cdev.14099

Murgiano, M., Motamedi, Y., & Vigliocco, G. (2021). Situating language in the real-world: The role of multimodal iconicity and indexicality. *Journal of Cognition*, *4*(1), Article 1. https://doi.org/10.5334/joc.113

Namy, L. L., Campbell, A. L., & Tomasello, M. (2004). The changing role of iconicity in non-verbal symbol learning: A U-shaped trajectory in the acquisition of arbitrary gestures. *Journal of Cognition and Development*, *5*(1), 37–57. https://doi.org/10.1207/s15327647jcd0501_3

Oldfield, R. C., & Wingfield, A. (1965). Response latencies in naming objects. *The Quarterly Journal of Experimental Psychology*, *17*(4), 273–281. https://doi.org/10.1080/17470216508416445

Özçalışkan, Ş, & Goldin-Meadow, S. (2005). Gesture is at the cutting edge of early language development. *Cognition*, *96*(3), B101–B113. https://doi.org/10.1016/j.cognition.2005.01.001

Pate, J. K., & Goldwater, S. (2015). Talkers account for listener and channel characteristics to communicate efficiently. *Journal of Memory and Language*, *78*, 1–17. https://doi.org/10.1016/j.jml.2014.10.003

Piantadosi, S. T., Tily, H., & Gibson, E. (2011). Word lengths are optimized for efficient communication. *Proceedings of the National Academy of Sciences*, *108*(9), 3526–3529. https://doi.org/10.1073/pnas.1012551108

Pluymaekers, M., Ernestus, M., & Baayen, R. H. (2005). Lexical frequency and acoustic reduction in spoken Dutch. *The Journal of the Acoustical Society of America*, *118*(4), 2561–2569. https://doi.org/10.1121/1.2011150

Priva, U. C. (2008). Using information content to predict phone deletion. In N. Abner & J. Bishop (Eds.), *Proceedings of the 27th West Coast conference on formal linguistics* (pp. 90–98). Cascadilla Proceedings Project. https://www.lingref.com/cpp/wccfl/27/abstract1820.html

Rowe, M. L., Ozçalişkan, S., & Goldin-Meadow, S. (2008). Learning words by hand: Gesture's role in predicting vocabulary development. *First Language*, *28*(2), 182–199. https://doi.org/10.1177/0142723707088310

Schröder, M., Charfuelan, M., Pammi, S., & Türk, O. (2008). *The MARY TTS entry in the Blizzard challenge 2008*.

Segal, J., & Newman, R. S. (2015). Infant preferences for structural and prosodic properties of infant-directed speech in the second year of life. *Infancy*, *20*(3), 339–351. https://doi.org/10.1111/infa.12077

Seyfarth, S. (2014). Word informativity influences acoustic duration: Effects of contextual predictability on lexical representation. *Cognition*, *133*(1), 140–155. https://doi.org/10.1016/j.cognition.2014.06.013

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, *27*(3), 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Shi, J., Gu, Y., & Vigliocco, G. (2022). Prosodic modulations in child-directed language and their impact on word learning. *Developmental Science*, *26*(4), Article e13357. https://doi.org/10.1111/desc.13357

Sjons, J., Hörberg, T., Bjerva, J., & Östling, R. (2017). *Articulation rate in Swedish child-directed speech increases as a function of the age of the child even when surprisal is controlled for*. https://doi.org/10.21437/Interspeech.2017-1052

Smith, N. J., & Levy, R. (2013). The effect of word predictability on reading time is logarithmic. *Cognition*, *128*(3), 302–319. https://doi.org/10.1016/j.cognition.2013.02.013

Soderstrom, M. (2007). Beyond babytalk: Re-evaluating the nature and content of speech input to preverbal infants. *Developmental Review*, *27*(4), 501–532. https://doi.org/10.1016/j.dr.2007.06.002

Stern, D. N., Spieker, S., Barnett, R. K., & MacKain, K. (1983). The prosody of maternal speech: Infant age and context related changes. *Journal of Child Language*, *10*(1), 1–15. https://doi.org/10.1017/S0305000900005092

Ter Bekke, M., Drijvers, L., & Holler, J. (2020, September 7). *The predictive potential of hand gestures during conversation: An investigation of the timing of gestures in relation to speech*. Gesture and Speech in Interaction Conference. https://pure.mpg.de/pubman/faces/ViewItemOverviewPage.jsp?itemId=item_3251942

Thiessen, E. D., Hill, E. A., & Saffran, J. R. (2005). Infant-directed speech facilitates word segmentation. *Infancy*, *7*(1), 53–71. https://doi.org/10.1207/s15327078in0701_5

Tippenhauer, N., Fourakis, E. R., Watson, D. G., & Lew-Williams, C. (2020). The scope of audience design in child-directed speech: Parents' tailoring of word lengths for adult versus child listeners. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *46*(11), 2163–2178. https://doi.org/10.1037/xlm0000939

Uther, M., Knoll, M. A., & Burnham, D. (2007). Do you speak E-NG-L-I-SH? A comparison of foreigner- and infant-directed speech. *Speech Communication*, *49*(1), 2–7. https://doi.org/10.1016/j.specom.2006.10.003

van Heuven, W. J. B., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). SUBTLEX-UK: A new and improved word frequency database for British English. *The Quarterly Journal of Experimental Psychology*, *67*(6), 1176–1190. https://doi.org/10.1080/17470218.2013.850521

van Schijndel, M., & Linzen, T. (2021). Single-stage prediction models do not explain the magnitude of syntactic disambiguation difficulty. *Cognitive Science*, *45*(6), Article e12988. https://doi.org/10.1111/cogs.12988

Vilà-Giménez, I., & Prieto, P. (2021). The value of non-referential gestures: A systematic review of their cognitive and linguistic effects in children's language development. *Children*, *8*(2), Article 148. https://doi.org/10.3390/children8020148

Vogt, S., & Kauschke, C. (2017). Observing iconic gestures enhances word learning in typically developing children and children with specific language impairment. *Journal of Child Language*, *44*(6), 1458–1484. https://doi.org/10.1017/S0305000916000647

Vosoughi, S., Roy, B., Frank, M. C., & Roy, D. (2010, May 11–14). *Effects of caregiver prosody on child language acquisition*. Proceedings of the 5th International Conference on Speech Prosody, Speech Prosody 2010, Chicago. https://doi.org/10.21437/SpeechProsody.2010-249

Wang, Y., Donnellan, E., & Vigliocco, G. (2023). How speech and representational gestures align in child-directed language: A corpus-based study. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *45*. https://escholarship.org/uc/item/4fz3×881

Watson, D. G., Arnold, J. E., & Tanenhaus, M. K. (2008). Tic Tac TOE: Effects of predictability and importance on acoustic prominence in language production. *Cognition*, *106*(3), 1548–1557. https://doi.org/10.1016/j.cognition.2007.06.009

Wonnacott, E., Brown, H., & Nation, K. (2017). Skewing the evidence: The effect of input structure on child and adult learning of lexically based patterns in an artificial language. *Journal of Memory and Language*, *95*, 36–48. https://doi.org/10.1016/j.jml.2017.01.005

Wu, Y. C., & Coulson, S. (2007). How iconic gestures enhance communication: An ERP study. *Brain and Language*, *101*(3), 234–245. https://doi.org/10.1016/j.bandl.2006.12.003

Yap, M. J., Tan, S. E., Pexman, P. M., & Hargreaves, I. S. (2011). Is more always better? Effects of semantic richness on lexical decision, speeded pronunciation, and semantic classification. *Psychonomic Bulletin & Review*, *18*(4), 742–750. https://doi.org/10.3758/s13423-011-0092-y

Zammit, M., & Schafer, G. (2011). Maternal label and gesture use affects acquisition of specific object names. *Journal of Child Language*, *38*(1), 201–221. https://doi.org/10.1017/S0305000909990328

Zangl, R., & Mills, D. L. (2007). Increased brain activity to infant-directed speech in 6- and 13-month-old infants. *Infancy*, *11*(1), 31–62. https://doi.org/10.1207/s15327078in1101_2

Zhang, Y., Frassinelli, D., Tuomainen, J., Skipper, J. I., & Vigliocco, G. (2021). More than words: Word predictability, prosody, gesture and mouth movements in natural language comprehension. *Proceedings of the Royal Society B: Biological Sciences*, *288*(1955), Article 20210500. https://doi.org/10.1098/rspb.2021.0500

Zipf, G. K. (1949). *Human behavior and the principle of least effort* (pp. xi, 573). Addison-Wesley Press.

Received October 10, 2022
Revision received February 22, 2024
Accepted March 7, 2024 ■