Title

Child-directed speech: the impact of variations in speaking rate on word learning

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 42(0)

Authors

Shi, Jinyu
Gu, Yan
Grzyb, Beata
et al.

Publication Date

2020

Peer reviewed

Child-directed speech: the impact of variations in speaking rate on word learning

Jinyu Shi (jessie.shi.17@ucl.ac.uk)
University College London, Psychology and Language Sciences, United Kingdom

Yan Gu (yan.gu@ucl.ac.uk)
University College London, Experimental Psychology, United Kingdom

Beata Grzyb (b.grzyb@ucl.ac.uk)
University College London, Experimental Psychology, United Kingdom

Gabriella Vigliocco (g.vigliocco@ucl.ac.uk)
University College London, Experimental Psychology, United Kingdom

Abstract

This study investigated how caregivers modulate their speaking rate according to children's lexical knowledge and the context of the interaction, and how such adjustments affect children's word learning. We studied a semi-naturalistic corpus where caregivers talked about different toys with their 3-4 years old children. The toys were known or unknown to the child, and present or absent from the environment. We found that caregivers talked about unknown toys with a slower speaking rate than known ones. When toys were absent, caregivers also tended to slow down for the toy's name, although they produced the whole utterance faster. Crucially, the results of a subsequent recognition task for children showed that caregivers' greater adjustment in speaking rate between known and unknown words predicted better immediate learning. Our findings suggest that caregivers modify their speaking rate in a helpful manner when the situation is more demanding, which assists children in word learning.

Keywords: child-directed speech; speaking rate; word learning; semi-naturalistic observation

Introduction

When talking to children, caregivers often use a specific speech register that is commonly referred to as child-directed speech (CDS). Compared to adult-directed speech (ADS), CDS is often cross-culturally characterized by a slower speaking rate, higher pitch, and greater pitch range (Fernald et al., 1989; Fernald & Simon, 1984; Soderstrom, 2007). To explain the prosodic modulations, Fernald (2000) put forward the idea that CDS is a type of hyper-speech (Lindblom, 1990), where speakers increase their signal quality based on their awareness of listeners' access to information that is independent of the signal. Since children could be considered as immature listeners, Fernald (2000) argued that caregivers modify their speech dynamically to facilitate children's word comprehension. Indeed, previous studies have shown that parents modulate their speech in ways that increase intelligibility. For example, when talking to children, caregivers have been shown to prolong their utterances (Cooper & Aslin, 1990), articulate the vowels more distinctly (Kuhl et al., 1997), and use a wider pitch range (Stern et al., 1983), all of which are considered to be characteristics of hyper-speech.

Amongst these characteristics, the current study focuses on the temporal quality (speaking rate, also known as articulation rate) of CDS, which was found to be related to toddler's word recognition as well as their vocabulary development. For instance, utilizing the preferential looking paradigm, Zangl et al. (2005) found that 12- to 31-month-old children recognized target words more accurately when the stimuli were presented with unaltered CDS compared to time-compressed CDS that was similar to ADS. Similar results were found in Song et al.'s (2010) experiment, where 19-month-old children identified a target picture faster and more accurately when they heard words with typical CDS than modified CDS that was twice as long (while keeping other prosodic features constant), suggesting that a slower speaking rate can enhance word recognition. Other research also showed that a slower speaking rate in CDS input to seven-months-olds was correlated to a better expressive vocabulary when they reached two (Raneri, 2015).

Despite the general interest in speaking rate, only a few studies have assessed whether, let alone how, caregivers adjust their internal prosodic pattern depending on children's knowledge or the context of the interaction. Prosody as a piece of signal-complementary information can provide contextual support to the children by highlighting focused words. Thus, caregivers should modify their speaking rate based on children's lexical knowledge, which could facilitate word learning. In particular, Han (2019) hypothesized that in word learning contexts, caregivers would slow down more when introducing words that are unknown to the child. For example, when native Dutch-speaking mothers read a storybook that contained five unfamiliar words and two familiar words to their 18 months old toddlers, they slowed down significantly for unknown words compared to known words both at word and utterance levels. Han (2019) further measured a speaking rate hyper-score for each mother by dividing the CDS score by the ADS score (ADS was collected as a control group), which captured how much the speaking rate differed between the same mother's CDS and ADS. She reported that compared to the general speaking rate hyper-score, the unknown utterance speaking rate hyper-score was a better predictor of children's receptive vocabulary at both 18 and 24 months. The results indicated that what aids children's acquisition of words is not the prosodic adjustment across utterances, but the particular modifications for unknown words. However, Han (2019) did not control for the frequency of the words, the referent positions in a sentence, or the number of mentions of the target referents in the discourse, despite these factors being able to significantly affect speaking rate (Gahl, 2008; Baker & Bradlow, 2009; Martin et al., 2016). Thus, the decrease in speaking rate might not be driven by caregivers' awareness of children's familiarity with different words, but by these confounding factors.

Besides children's familiarity with the words, the context of the interaction (i.e., situational accessibility of the referents) may also affect the speaking rate of words in CDS. Previous research mostly investigated CDS speaking rate when the referents of the target words were present in front of the caregiver and the child (Fernald & Mazzie, 1991; Han, 2019; Raneri, 2015), but in real life, a sizable portion of the input to children refers to referents absent from the immediate environment (i.e., displaced) (Veneziano, 2001). Thus, besides learning the target referents in a situated context, children also need to be able to access the association between the labels and the referents in the displaced context. Research in information structure has argued that in a conversation, referents constantly change between three statuses through activation or deactivation: given (completely active in a person's focus of consciousness), new (not active at all), or accessible (in a person's peripheral consciousness but is not what being focused) (Chafe, 1987). One way for the referent to be accessible for the listener is for it to simply be a part of the physical environment, which is termed as situationally accessible (Lambrecht, 1994). Hence, compared to situated contexts, the referents would be less accessible to the children when the conversation happened in a displaced context. This means that more cognitive effort would be required for children to activate or build the association between the referent labels with the actual objects in displaced contexts. As studies with adults showed that English speakers tend to use accented expressions for referents that are less accessible to the listener (Arnold, 2008), it is likely that caregivers would also slow down their speaking rate in displaced contexts to compensate for the increase in children's cognitive load.

However, conversing about a referent in displaced contexts also means that, for children, as the visual information about the referents is no longer available, most of the information they can get on the referent comes from the caregivers' speech. Thus, caregivers might show a tendency to provide verbally more information about the referent, so that children can conceptualize or retrieve the referent sooner and better. Thus, compared to situated contexts, caregivers may try to fit in more information within the same amount of time when the referents are absent from the environment, which may result in a faster speaking rate. This provides an opportunity to test how information efficiency interacts with information accessibility in modulating speaking rate. Another possible reason why caregivers increase their speaking rate in displaced contexts could simply be children's lack of attention when the objects are not in front of them. If children were clearly impatient and not paying attention to the interaction, caregivers may rush through the information related to absent objects, in the hope that the child will get some of it. Given that no previous study has so far explored the effect of displacement on CDS, this study will be the first to examine how caregivers modify their speaking rate in a displaced context.

Previous studies have also looked at whether word acquisition is linked with either the speaking rate across utterances in CDS (Raneri, 2015) or how different the mother's CDS is to ADS (Han, 2019; Kalashnikova & Burnham, 2018). However, there is generally a lack of research investigating how the internal speaking rate modulations relate to word learning. For interactions that use CDS relatively consistently, what matters for children is probably not how different the speaking rate is in CDS compared to ADS, but to what extent the caregivers mark the information that children need to learn (i.e., the unknown words). The possible amendment in speaking rate within CDS could highlight the unknown words and provide prosodic cues to the children. Thus, if children actually make use of these cues when acquiring words, we would expect a correlation between the degree of speaking rate modulations on unknown words within CDS and children's word learning outcomes. Such a relation has not been investigated before, despite its potential interest.

Thus, the present study investigates whether (1) English speaking caregivers adjust their speaking rate in situations that are more demanding to the children (i.e., when the referents are unknown to the child or absent from the environment), and (2) the speaking rate of unknown words/utterances, or the degree of speaking rate modification between known and unknown words/ utterances (speaking rate ratio between known and unknown) predict children's outcome of recognizing unknown words they previously encountered and their general lexical development.

To address (1), we used data from a semi-naturalistic corpus (the ECOLANG corpus, Vigliocco et al., in prep) that introduced learning and displaced contexts. The caregivers were asked to interact with their child by talking about the toys that were either known or unknown to the child under two conditions: the target toys were present or absent in the context. To address (2) we tested children's vocabulary size, as well as how much they learned after the interaction.

Method

Participants

This study included 30 dyads (caregiver-child, all mothers, 14 girls) from the ECOLANG corpus (Vigliocco et al., in prep). Participants were all native English speakers recruited from the wider London area. The age of the children ranged from 3 to 4.33 years (mean = 3.53 years). The mean age of mothers was 38.57 years, ranging from 34 to 45 years old. All but one mother had received a bachelor's degree or above (the mean educational level was 4.43, where 1 = GCSE, 6 = doctoral degree). The study obtained ethical approval from UCL.

Materials

Four sets of toys were used in the task, comprising of animals, tools, foods, and musical instruments. For each dyad, six toys per set (24 toys in total) were included, where three of them were known to the child and three were unknown. Prior to experiment days, parents were sent a wordlist in which they would be asked to indicate which toys their children had already known (label and concept) without checking back with the children. The assignment of toys to familiarity condition was based on the answers. Each toy was used for roughly an equal number of dyads.

The British Picture Vocabulary Scale 3rd edition (BPVS3) was used to test the children's vocabulary size (Dunn et al., 2009). A recognition test that asked the child to identify the unknown objects they just talked about was registered for each child to test how much they have learned from the interaction.

Procedure

The interactions took place at participants' homes. The BPVS3 was administered before the interaction began. During the interaction, the caregiver and the child were sitting at 90 degrees from each other around a table. Two video cameras (one focusing on the caregiver, one focusing on the child and the interaction space) recorded the interaction; the speech of the caregiver was also recorded using a clip-on microphone via Audacity.

During the sessions, the caregiver was instructed to talk about the known and unknown toys in each set in a natural way (i.e., how they usually interact with their children) under either toy-present or toy-absent condition. The two conditions were counterbalanced across participants. In the toy-present condition, the experimenter brought six toys from one set (e.g., animal) to the table and left the room. The caregiver and the child talked about and interacted with the toys for 3-4 minutes. The experimenter then reentered the room, asked the child to help tidy up the toys, and left the room together with the toys. In the toy-absent condition, the experimenter asked the caregiver to continue to talk about the toys they just played with (when toy-present was first) or the toys that were about to come (when toy-absent was first) for 3-4 minutes. Labels of the toys would be provided as a reminder for the caregivers. This process would be repeated for all four categories, resulting in 8 sessions in total for each dyad. Conditions and sessions were counterbalanced in sequences. The whole recording lasted approximately 25-35 minutes.

After the interaction, children received a recognition test presented in E-prime. In each trial, two pictures, either both unknown toys or both known toys used in the interaction, were presented at the left and right side on the computer screen and a female native British English speaker would articulate a question such as 'Can you help me to find the [pomegranate], where is the [pomegranate]?'. The child was asked to point to which one of the two pictures displayed on the screen matched with the word they heard. The recognition test contained 24 test trials (identification of the unknown words) and 4 control trials (identification of the known words).

Coding

Speech was transcribed manually by utterances utilizing Praat (Boersma & Weenink, 2019). An utterance corresponds to an event or predicate unit (see Berman & Slobin, 1994; Vigliocco et al., 2019). Each audio clip and its transcription were also automatically segmented and aligned on a word level via the Munich Automatic Segmentation System (MAUS, https://clarin.phonetik.uni-muenchen.de/BASWebServices/interface/WebMAUSGeneral), and the segmentations were manually checked and improved in accuracy. The target referents were coded for conditions: familiarity level (known vs. unknown), and object present vs. absent. Additionally, unknown words are generally less frequent than known words, and high-frequency words are usually produced faster than words that are less frequent (Gahl, 2008). Hence, we also extracted the frequency for each target word from the SUBTLEX-UK dataset (van Heuven et al., 2014) to account for caregivers' experiences with the words.

The speaking rate was measured by the average number of syllables per second. Pauses longer than 200 ms and occurred within utterances were subtracted from the durations (Han, 2019).

To measure the degree of modification in speaking rate between known and unknown words, we divided the mean speaking rate for the known words by that of the unknown words for each caregiver and obtained the speaking rate ratio at the word level. The same calculation was done for the 'known' and 'unknown' utterances that contain the target words to obtain the utterance level speaking rate ratio.

Children's recognition test scores were calculated by dividing the number of testing trials the child answered correctly by the total number of testing trials (i.e., the proportion of trials that the child answered correctly). We also coded caregivers' average number of mentions for the unknown words as a control variable because children could probably learn better when they heard the words multiple times. The standardized score for BPVS3 was calculated for each child based on the raw score.

Results

We used linear mixed effect models in the R environment (R Core Team, 2020) to assess which factors influence caregivers' speaking rate. The model included the presence or absence of the object and the familiarity of the word (known/unknown) as centered fixed effects. We included random intercept and random slope for participants and words. The dependent variable was either caregivers' speaking rate of the target utterances or referents depending on the level of analysis (utterance or word).

To eliminate the effect of confounding factors that may affect CDS and speaking rate in general, we entered several the following factors as control variables for the utterance level analysis, namely the age of the children (Narayan & McDermott, 2016) and whether the section that the referent occurred in was the first (new) or the second (repeated) section (e.g., in present-absent conditions, the present session was always new while the absent was repeated and vice versa). Since repeated information is usually articulated faster (Arnold, 2008), the speaking rate of the target words might be affected depending on what section they were in. For the word level analysis, we also included the number of mentioning of the target words, word position in an utterance (initial vs. medial vs. final vs. isolation), and the frequency of the words as control variables that have been shown to influence speaking rate.

In total, 4693 target utterances (2236 with known words) and were included in the utterance analysis. The results of a regression model (Table 1) showed that there was a significant main effect of both familiarity and condition (toys being present or absent), while controlling for age and session order (neither affect the utterance speaking rate significantly). There was no significant interaction between familiarity and condition (β=-0.102, *SE*=0.092, *t*= 1.113, *p*=0.27), and including this interaction term did not improve model fit (χ^2=1.25, *p*=0.26), so it was dropped in the final model. The results suggest that the speaking rate was (1) slower when the utterance contains unknown words in both conditions but (2) faster in toy-absent condition compared to toy-present condition.

Table 1: Summary of the model for utterance speaking rate

	Estimate	Std. Error	t	p
(Intercept)	4.80	0.092	51.79	<0.001
PresentAbs	0.15	0.068	2.22	0.035
Familiarity	-0.30	0.090	-3.30	0.0016
Age	0.05	0.23	-0.20	0.84
NewRepeated	-0.002	0.067	-0.03	0.97

For the word level analysis, a total of 5327 (2562 known words) instances of target referent was included. The regression model (Table 2) showed that besides age, all control variables were highly significant in influencing the speaking rate. When these control variables were taken into account, we found a significant main effect of familiarity and a marginal significant main effect of the absent vs. present condition. As Figure 2 shows, the speaking rate was consistently slower when the words were unknown and when the objects were absent. The interaction between familiarity and condition was again not significant (β = -0.036, *SE* = 0.074, *t* = -0.484, *p* = 0.63) and did not improve model fit (χ^2 = 0.23, *p* = 0.63), so it was not included in the model. The results suggested that (1) unknown words were produced slower than known words in both present and absent condition; (2) speaking rate of target referents tended to be slower when the toys were absent compared to when they were present.

Table 2: Summary of the model for target word speaking rate (syllables/second)

	Estimate	Std. Error	t	p
(Intercept)	3.85	0.13	30.52	<0.001
PresentAbs	0.094	0.052	1.79	0.081
Familiarity	-0.19	0.084	-2.22	0.032
Age	0.016	0.014	1.14	0.27
Frequency	-0.39	0.14	-2.68	0.0086
Num.Mention	0.072	0.0071	10.14	<0.001
Initial	0.85	0.083	10.20	<0.001
Medial	1.05	0.036	29.30	<0.001
Isolation	-0.42	0.059	-7.14	<0.001
NewRepeated	0.12	0.048	2.45	0.02

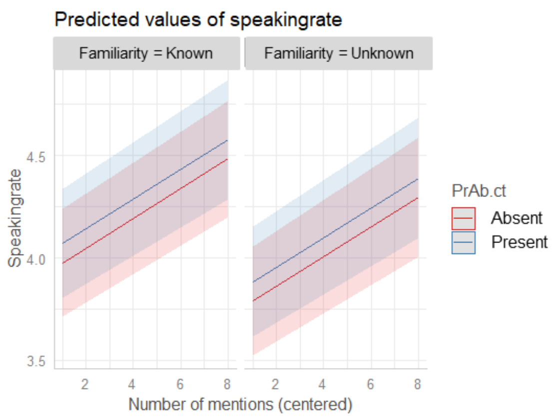

Figure 1: Predicted values of speaking rate based on the model, showing how the speaking rate increased as the number of mentions increased for known and unknown words in absent and present conditions.

We then looked at the effect of caregivers' speaking rate on children's learning of the word labels and their long-term lexical development. We used multiple linear regressions with caregivers' speaking rate of unknown words or speaking rate ratio as separate independent variable. The dependent variable was the children's recognition results or BPVS3 results.

For children's learning of the referents, we found that they answered on average a proportion of 0.81 test trials correctly (*SD*=0.14). Although recognition test results were not significantly correlated with the raw speaking rate of the unknown words or utterance measurements (all *p* >.05), there was a significant correlation between caregivers' word articulation rate ratio and children's recognition learning result (β = 0.4, *SE* = 0.16, t = 2.61, p = 0.014) (Figure 2). Such an effect still held when the average number of mentions of the unknown referents and children's age were controlled for (β = 0.39, *SE* = 0.16, t = 2.38, p = 0.025), or their vocabulary size was controlled for (β = 0.52, *SE* = 0.15, t = 3.48, p = 0.0019). The analyses indicated that instead of having a slow speaking rate in general, it is caregivers' greater adjustment between known and unknown words that predicted better immediate learning for children.

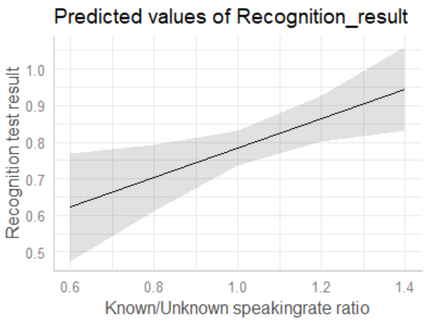

Figure 2: The model's prediction of recognition test results based on the speaking rate ratio

For the general vocabulary size, children received a score of 107.33 on average (*SD*= 10.76). There was no significant correlation found between any of the independent variables and the BPVS3 results (all p >.05).

Discussion

The current study investigated how caregivers' speaking rate varied according to different contextual factors in the input to 3- to 4-year-old children and whether such variation could potentially support word learning.

Firstly, we looked at whether the speaking rate of the caregiver differed when they were talking about toys that were unknown to the child compared to those that were known. For the utterance level analysis, we controlled for the age of the child and session repetition, which did not affect utterance speaking rate significantly. This is not completely unexpected as caregivers might often repeat the target labels but rarely repeat the whole utterance. For the word level analysis, we further controlled for factors that could particularly affect speaking rate of individual words, such as word position, the number of mentions, and word frequency. As expected, all these factors affected speaking rate significantly. Beyond that, we found that the speaking rate was consistently slower for unknown referents compared to known words at both word and utterance levels, in line with our predictions and the study by Han (2019). The results suggested that using a slow speaking rate, caregivers highlight the words that they consider as unknown and therefore more difficult to the children, which supports the idea that caregivers dynamically modify their speech depending on their perception of children's lexical knowledge. Additionally, since the familiarity effect was found when we controlled for the frequency of the words, our results indicated that the decrease in speaking rate is not only a speaker-oriented effect (i.e., speakers are slower for less frequent words) but crucially a listener-oriented effect (i.e., expectations about children's lexical knowledge).

Besides the effect of children's familiarity on the speaking rate of target words, we also looked at how the situational accessibility of the referents might affect caregivers' speaking rate. We found that at the utterance level, caregivers have a faster speaking rate when the corresponding toys were absent compared to when they were present. In contrast, we observed a reversed pattern in the word level analysis: the caregivers tended to slow down for the target referents in displaced contexts compared to situated contexts, though these differences were marginally significant.

We propose that the differences in prosodic patterns can be explained by caregivers' desire to be informative and compensate for the low accessibility of the referents (that is caused by the displaced contexts). In displaced contexts, referents can only be accessed via the language input. Therefore, to help children to retrieve (or build) an association between a label and a referent, caregivers would try to provide more verbal information in a similar amount of time (as the child could lose interest otherwise). However, they also use strategies (such as use of prosodic cues) to compensate for children's cognitive load, resulting in a tradeoff between speaking rate and informativeness. Our results suggest that caregivers achieve both functions by modulating their speech differently on distinct levels of speech. At the utterance level, caregivers increased their speaking rate in displaced contexts so that they could relay more information in the same amount of time. At the word level, it is only about the particular label of the referents. Thus, to potentially help children to process the words, caregivers would dynamically adjust their speaking rate of the label based on how accessible the referents are to the children.

To further explore what may cause the opposite patterns of results in speaking rate between the word and utterance levels, we conducted a post-hoc analysis examining how the speaking rate varied when looking at the utterances without the target word. Take the utterance 'mommy really likes to play drum' as an example: the target label in this utterance is 'drum', and the utterance without target word is 'mommy really likes to play'. If caregivers try to be as informative as possible, then we would expect an increase in speaking rate for displaced contexts when the caregivers were saying the utterance without target word. Indeed, the results of a regression analysis showed that caregivers articulated the rest of the utterance significantly faster in the absent condition (β = 0.17, *SE* = 0.052, t = 3.27, p = 0.032) and when related to known words (β = -0.30, *SE* = 0.062, t = -4.78, p < 0.001). The results suggested that the increase in speaking rate of utterance for absent condition is likely to be driven by the rest of the sentence, supporting our speculation that caregivers may introduce referent related information faster to be more informative to children. Additionally, the speaking rate was also slower for those related to unknown words, which is consistent with our previous results. Together with the previous results that show the speaking rate for target words tended to be slower in absent condition, our findings suggest that caregivers increased their speaking rate to be informative while slowing down for the target referent to make it acoustically salient.

Furthermore, we investigated the relationship between speaking rate and children's word learning. Despite the fact that neither measurements on utterances nor the raw speaking rate of words were correlated with the recognition results for unknown words, we detected a significant correlation between the speaking rate ratio and recognition test results. This indicated that the greater the adjustment the caregiver made to her speaking rate for unknown words relative to the known words, the better the immediate learning results are for children. Unlike the previous belief that a slower speaking rate in general helps children to learn words, our finding suggests that what affects children's immediate word learning is how much the novel target words were marked in the speech. When the caregiver is speaking with a consistent and invariant rate, no matter how fast or slow they are, the target referents that children need to learn are not prominently highlighted. Thus, children may not be able to utilize the prosodic pattern as a cue to word learning. On the contrary, when the unknown words are articulated much slower than their known counterpart, they might attract children's attention and even provide more linguistic information about the word itself, resulting in better immediate learning outcomes.

However, we did not find evidence supporting that speaking rate of CDS is linked with 3-4 year-old children's long-term lexical development. This is not completely unexpected because speaking rate is only one aspect of prosody, let alone other linguistic factors like the gestures used (Vigliocco et al., 2019) and environmental factors like schooling that also strongly affect lexical development. Past research has shown that a slower articulation rate of CDS correlated with children's larger vocabulary size (Raneri, 2015). The inconsistency between our result and earlier findings might result from differences in measurements, since we focused on word learning context with unknown words. Also, as the age of the children in our study was above three, it is possible that they do not rely as much on the prosodic cues compared to preverbal infants. Additionally, the current correlation study between vocabulary and speaking rate was based on a small sample size of 30 children. Future studies could look at caregivers' prosodic modification on unknown words and how it correlates to children's word learning with a larger and age-spread sample.

Additionally, it would be interesting to investigate whether a similar pattern would appear in other aspects of prosody, such as pitch and intensity. In fact, having a higher pitch and intensity are also essential characteristics of CDS (Soderstrom, 2007) and were often used to highlight contextually new or focused information (Bortfeld & Morgan, 2010). For example, Han (2019) examined whether the mean pitch and pitch range in CDS differed depending on children's familiarity with the target referents and found mixed results. For Mandarin speakers, caregivers have an increased utterance mean pitch and exaggerated pitch range for unknown words, but Dutch caregivers raised their mean pitch for known words instead. Therefore, exploration of how our English caregivers would adjust other prosodic cues in their speech based on children's lexical knowledge and the accessibility of the information, and how such adjustments might affect children's word learning would also contribute to this area.

Conclusion

Previous researchers have looked at how the prosodic modulations in CDS could facilitate word learning, but few have studied the internal prosodic pattern of CDS in word learning context and its effect on children's acquisition of words. The current study found that caregivers slow down their speaking rate when talking about unknown referents with their child, suggesting that caregivers dynamically modify their speech with didactic intent to facilitate word learning. We also discovered that although caregivers increase their speaking rate of utterances in general, they still slow down for the particular words when the target referents were absent from the physical environment (i.e., less accessible), indicating that parents do adjust their speech based on their assumption of the children's cognitive load. To our best knowledge, this is the first study comparing the CDS speaking rate between conditions where objects were present and absent. Together, both findings support the idea that speakers are aware of the listener's mental model and are constantly amending their speech based on such awareness. Moreover, our results showed that contrary to common belief, children's immediate word learning outcome did not correlate with the raw speaking rate measurements, but with the degree of modulation between the known and unknown words, which implied that children actually make use of the prosodic cues provided when learning. These findings provide a more comprehensive understanding of the prosodic pattern inside CDS and how it links to children's lexical development.

Acknowledgement

The work reported here was supported by a ERC Advanced Grant (743035) to GV, and a NWO Rubicon Grant (019.182SG.023) to YG.

References

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

Baker, R. E., & Bradlow, A. R. (2009). Variability in Word Duration as a Function of Probability, Speech Style, and Prosody. *Language and Speech*, *52*(4), 391–413.

Berman, R. A., & Slobin, D. I. (1994). *Relating events in narrative: A crosslinguistic developmental study*. Hillsdale, N.J. : L. Erlbaum Associates.

Boersma, P., & Weenink, D. (2019). *Praat: Doing phonetics by compute* (Version 6.1.08) [Computer software].

Chafe, W. (1987). Cognitive constraints on information flow. *Coherence and Grounding in Discourse*, *11*, 21–51.

Cooper, R. P., & Aslin, R. N. (1990). Preference for Infant-Directed Speech in the First Month after Birth. *Child Development*, *61*(5), 1584–1595. JSTOR.

Dunn, D. M., Dunn, L. M., National Foundation for Educational Research in England and Wales, & GL Assessment (Firm). (2009). *The British picture vocabulary scale.* GL Assessment.

Fernald, A. (2000). Speech to Infants as Hyperspeech: Knowledge-Driven Processes in Early Word Recognition. *Phonetica*, *57*(2–4), 242–254.

Fernald, A., & Mazzie, C. (1991). Prosody and focus in speech to infants and adults. *Developmental Psychology, 27*(2), 209–221.

Fernald, A., & Simon, T. (1984). Expanded intonation contours in mothers' speech to newborns. *Developmental Psychology*, *20*(1), 104–113.

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. Cambridge Core.

Gahl, S. (2008). 'Time' and 'Thyme' Are Not Homophones: The Effect of Lemma Frequency on Word Durations in Spontaneous Speech. *Language*, *84*(3), 474–496. JSTOR.

Han, M. (2019). *The Role of Prosodic Input in Word Learning: A Cross-linguistic Investigation of Dutch and Mandarin Chinese Infant-directed Speech*. LOT Publisher: Amsterdam.

Hockett, C. F. (1960). The Origin of Speech. *Scientific American*, 203(3), 88–97. JSTOR.

Kalashnikova, M., & Burnham, D. (2018). Infant-directed speech from seven to nineteen months has similar acoustic properties but different functions. *Journal of Child Language, 45*(5), 1035–1053. Cambridge Core.

Kuhl, P. K., Andruski, J. E., Chistovich, I. A., & Chistovich, L. A. (1997). Cross-language analysis of phonetic units in language addressed to infants. *Science*, *277*(5326), 684–686.

Lambrecht, K. (1994). *Information Structure and Sentence Form: Topic, Focus, and the Mental Representations of Discourse Referents*. Cambridge University Press; Cambridge Core.

Lindblom, B. (1990). Explaining Phonetic Variation: A Sketch of the H&H Theory. In W. J. Hardcastle & A. Marchal (Eds.), *Speech Production and Speech Modelling* (pp. 403–439). Springer Netherlands.

Martin, A., Igarashi, Y., Jincho, N., & Mazuka, R. (2016). Utterances in infant-directed speech are shorter, not slower. *Cognition*, *156*, 52–59.

Narayan, C. R., & McDermott, L. C. (2016). Speech rate and pitch characteristics of infant-directed speech: Longitudinal and cross-linguistic observations. *The Journal of the Acoustical Society of America*, *139*(3), 1272–1281.

Raneri, D. P. (2015). *Infant-directed speech: Maternal pitch variability, rate of speech, and child language outcomes*. University of Maryland, College Park.

R Core Team. (2020). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing. http://www.R-project.org/

Soderstrom, M. (2007). Beyond babytalk: Re-evaluating the nature and content of speech input to preverbal infants. *Developmental Review*, *27*(4), 501–532.

Song, J. Y., Demuth, K., & Morgan, J. (2010). Effects of the acoustic properties of infant-directed speech on infant word recognition. *The Journal of the Acoustical Society of America*, *128*(1), 389–400.

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.

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.

Veneziano, E. (2001). Displacement and informativeness in child-directed talk. *First Language, 21*(63), 323–356.

Vigliocco, G., Murgiano, M., Motamedi, Y., Wonnacott, E., Marshall, C. R., Milán-Maillo, I., & Perniss, P. (2019). Onomatopoeias, gestures, actions and words: How do caregivers use multimodal cues in their communication to children? In A. K. Goel, C. M. Seifert, & C. Freksa (Eds.), *Proceedings of the 41th Annual Meeting of the Cognitive Science Society, CogSci 2019: Creativity + Cognition + Computation, Montreal, Canada, July 24-27, 2019* (pp. 1171–1177). cognitivesciencesociety.org. https://mindmodeling.org/cogsci2019/papers/0215/index.html

Vigliocco, G., Gu, Y., Motamedi, Y., Grzyb, B., Brieke, R., Murgiano, M., Perniss, P. (in prep). The ECOLANG corpus of dyadic interactions between caregivers and their 2-4 year-old child and between two adults.

Zangl, R., Klarman, L., Thal, D., Fernald, A., & Bates, E. (2005). Dynamics of Word Comprehension in Infancy: Developments in Timing, Accuracy, and Resistance to Acoustic Degradation. *Journal of Cognition and Development*, *6*(2), 179–208.