CrossMark click for updates

A multimodal corpus of speech to infant and adult listeners

Elizabeth K. Johnson[a)]
University of Toronto Mississauga, 3359 Mississauga Road, Mississauga, Ontario L5L 1C6, Canada
elizabeth.johnson@utoronto.ca

Mybeth Lahey
Max Planck Institute for Psycholinguistics, POB 310, 6500 AH Nijmegen, The Netherlands
mybeth.lahey@mpi.nl

Mirjam Ernestus
Centre for Language Studies, Radboud University Nijmegen, Postbus 9103, 6500 HD Nijmegen, The Netherlands
m.ernestus@let.ru.nl

Anne Cutler
The MARCS Institute, University of Western Sydney, Penrith South, New South Wales 2751, Australia
A.Cutler@uws.edu.au

Abstract: An audio and video corpus of speech addressed to 28 11-month-olds is described. The corpus allows comparisons between adult speech directed toward infants, familiar adults, and unfamiliar adult addressees as well as of caregivers' word teaching strategies across word classes. Summary data show that infant-directed speech differed more from speech to unfamiliar than familiar adults, that word teaching strategies for nominals versus verbs and adjectives differed, that mothers mostly addressed infants with multi-word utterances, and that infants' vocabulary size was unrelated to speech rate, but correlated positively with predominance of continuous caregiver speech (not of isolated words) in the input.

© 2013 Acoustical Society of America
PACS numbers: 43.71.Ft [SGS]
Date Received: August 1, 2013 **Date Accepted:** October 14, 2013

1. Introduction

Among the many remarkable aspects of language acquisition is the ease and rapidity with which young children acquire a vocabulary. Vocabulary learning begins well before its results can be turned into useful conversations; the earliest produced words appear generally very late in the first year of life, but evidence for recognition of spoken forms, including the child's own name,[1] names of body parts,[2,3] and the names of caregivers,[4] has been found many months earlier.

The language input that infants are exposed to is, of course, crucial in this acquisition process. Although some have argued that the production of isolated words plays a crucial role in early language development,[5,6] the available evidence overwhelmingly suggests that speech directed to infants consists largely of multi-word utterances,[7,8] forcing young language learners to extract word forms from surrounding speech context to store them in memory for subsequent recognition. A lot is known about the characteristics of input to infants in the first year of life. For example, in

[a)]Author to whom correspondence should be addressed.

 © 2013 Acoustical Society of America

Redistribution subject to ASA license or copyright; see http://acousticalsociety.org/content/terms. Download to IP: 195.169.108.229 On: Tue, 26 Nov 2013 11:26:17

comparison with speech directed to adult listeners, speech to infant listeners generally has a slower speaking rate, exaggerated prosody, shorter utterances, greater expressive affect, and an expanded vowel space (e.g., Refs. 8–10). Importantly, many of these properties have been shown to correlate with infant success at word-form perception and learning,[11–14] crucially linking the nature of the language input to subsequent language learning success.

Laboratory studies of infant-directed speech, from which this picture has arisen, have typically contrasted productions by a mother addressing her own infant versus the same mother addressing an unfamiliar adult (often the experimenter). This approach has enabled researchers to collect well-controlled clear audio recordings of infant- and adult-directed speech. However, the speech the mother directs to an unfamiliar experimenter encountered in a laboratory setting may not be representative of the speech the mother would direct to other more familiar adults commonly encountered in the child's everyday environment. In addition, the child's typical environment may also often include multiple simultaneous interlocutors, forcing the child to separate overlapping speech streams to extract useful linguistic information from spoken input. Some of the properties of infant-directed speech mentioned in the preceding text have also been studied in more naturalistic recordings made in the home, whereby most of these naturalistic recordings with children under a year have involved dense recordings with a single child (e.g., Ref. 8). Such dense corpora are highly useful but do not readily allow researchers to assess generalizability or to examine the relationship between different types of caregiver input and children's developing language skills. That is, existing corpora tend to be naturalistic but less generalizable or generalizable but less naturalistic.

In this report, we describe an extensive new corpus that (a) involves recordings of input provided (to 28 Dutch-learning infants) at the very onset of the natural word-teaching period, namely just before the first birthday; (b) involves speech from multiple caregivers interacting with the infant together as well as speech between caregiver(s) and experimenter; (c) includes a video record from two separate angles to avoid loss of relevant data; and (d) also compares teaching of words from different word classes (noun, proper name, verb, adjective).

Data from this corpus will be generalizable but also naturalistic. Such an extensive corpus consisting of both semi-structured word teaching activities and free interactions allows a wide variety of differing questions to be addressed, including, but not limited to, questions about the acoustics of the speech input, the relation of speech characteristics to eye contact, potential differences in adult-to-child interactions during explicit word teaching tasks versus free interaction, the consistency of speech acoustics across talkers and across interlocutor groups, and the variation of speech acoustics as a function of word class. The primary purpose of this summary overview is to announce the public availability of these data to the field of infant speech perception research. In addition, we show how the data can help to answer three specific questions of general interest to infant speech perception researchers: How does mean length of utterance vary across speech directed to infants, to familiar adults, and to unfamiliar adults? Are words of different word classes taught in the same way? Have children with larger productive vocabularies been exposed to more isolated word tokens? Answers to these questions, and to many more which the corpus will enable researchers to address, will help to improve our understanding of the crucial relationship between speech input and language learning.

2. Data acquisition

The corpus comprises 33 h of spoken language interaction in 65 play sessions recorded between September 2005 and August 2007 at the Max Planck Institute Baby Research Centre in Nijmegen, The Netherlands. Each recording involved an infant, the infant's mother or father, and an experimenter, with most recordings also involving an additional caregiver. Both free play and word teaching activities were recorded as well as

Redistribution subject to ASA license or copyright; see http://acousticalsociety.org/content/terms. Download to IP: 195.169.108.229 On: Tue, 26 Nov 2013 11:26:17

adult-to-adult conversations. All interactions were audio-taped using radio-controlled microphones clipped to the two caregivers and videotaped from perpendicular angles (see Fig. 1). To assess vocabulary development, the Dutch version of the MacArthur Bates Child Developmental Inventory (N-CDI, short form[15]) was collected at the time of the first session and a follow-up N-CDI was collected approximately 1 yr later.

2.1 Participants

Twenty-eight triads were recorded; each triad included a typically developing infant with reportedly normal hearing (20 males; 8 females), a parent of the infant, and an additional primary caregiver (usually the other parent or grandmother). All participants lived in the Nijmegen region at the time of the recording sessions, and all but one were native Dutch speakers. Parental region of origin, education level, and approximate age were noted.

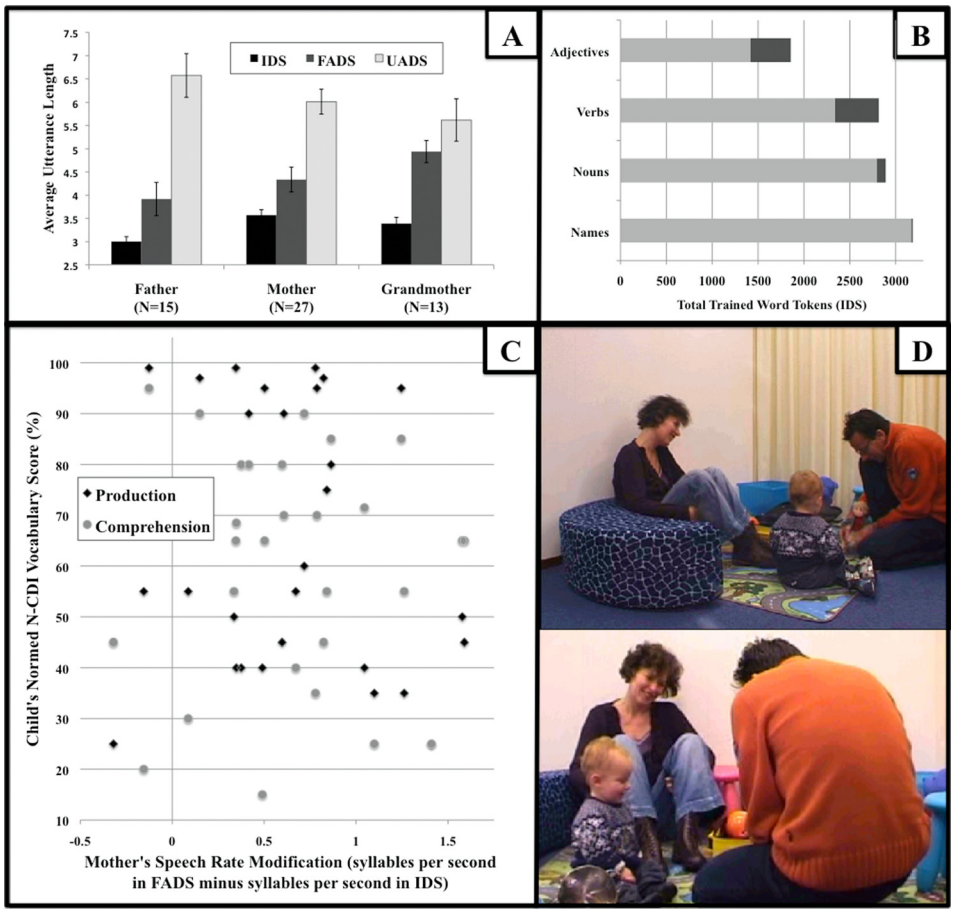

Fig. 1. (Color online) (A) Average utterance length in words as a function of speech register (infant-directed speech, familiar-adult-directed speech, and unfamiliar-adult-directed speech). Error bars indicate standard error. (B) Total number of infant-directed target words produced by caregivers. Base forms are indicated in light gray (e.g., "knikken" for "knikken"), inflected forms are indicated in dark gray and included comparatives (e.g., "hariger" for adjective "harig"), diminutives ("koffertje" for noun "koffer"), conjugations (e.g., "stamp" for verb "stampen"), possessives (e.g., "Roemers" for proper name "Roemer"), etc. (C) Mother's degree of speech rate modification plotted as a function of children's productive vocabulary scores at the time of the recording. (D) Example screenshot of two simultaneous camera angles recorded during each session. In general, the gaze direction of all participants can be retrieved from the video record.

Redistribution subject to ASA license or copyright; see http://acousticalsociety.org/content/terms. Download to IP: 195.169.108.229 On: Tue, 26 Nov 2013 11:26:17

Infant participants were approximately 11 months old at the time of the first recording session (range = 11:0–12:6; M = 351 days; SD = 11:7). Half of the participants ($N = 14$) were recorded on two further occasions within 20 days after the first session, but not all participants were able to commit to three sessions. For nine participants, the corpus contains only one session and for five participants, two sessions.

2.2 Recording sessions

Sessions were recorded in a quiet room at the Baby Research Centre in Nijmegen and ranged in duration from 21 to 40 min (M = 30.5 min). Recording of each session began with a five-min free play period, followed by 12 min of word teaching. All recording sessions ended after various activities designed to elicit unscripted adult-to-adult interactions. Thus transcribed interactions involved ample free interactions and discussions in addition to the targeted word teaching activities.

In the first session, caregivers were given four words to teach their child: a proper name (*Zanthe, Mirre, Roemer,* or *Tigo*), a common noun (*cactus,* cactus; *koffer,* suitcase; *zadel,* saddle; or *masker,* mask), an adjective (*glanzend,* shiny; *krullend,* curly; *puntig,* pointy; or *harig,* hairy), and a verb (*stampen,* to stomp; *krabben,* to scratch; *buigen,* to bow; or *knikken,* to nod). Each triad was assigned one of each of the four word types to teach their child with different triads given different combinations of words. All these bisyllabic target words had a trochaic stress pattern and were deemed unlikely to be known by the average 11-month-olds. To facilitate word teaching, props were provided, including a doll with the to-be-taught name, and relevant objects (e.g., a toy cactus, or a shiny colander). If caregivers returned for a second session, they were given four new words to teach their child. For example, if a child learned *cactus, Tigo, glanzend,* and *stampen* on the first visit, they might be taught *koffer, Mirre, harig,* and *krabben* on the second visit. Immediately following each Day 1 or 2 word teaching task (and before the recording ended), caregivers spent some time filling out paperwork together. Much of the read speech in this corpus resulted from the caregivers reading these forms to each other. To elicit adult-to-adult conversation, the caregivers were encouraged to work together on answering the questions on the forms. Spontaneous conversations between the adults also occurred during the recording sessions. The experimenter returned toward the end of the session to chat with the caregivers. Recorded interaction with the experimenter included free conversation as well as discussion of paperwork and word teaching strategies. The first two visits always involved two caregivers plus an experimenter; the third visit involved only one caregiver. In the third visit, the caregiver reviewed the words taught during the first two sessions.

At the outset of all sessions, caregivers were asked to spend equal time teaching each word type (noun, verb, adjective, name). They were not told how to teach the words (e.g., they were not told to put target words in complete sentences or to avoid diminutive forms); they were, however, told that their children would be tested later on their knowledge of the words. First and second visits thus finished with a Headturn Preference Experiment[16] to assess infants' recognition of the four word forms taught during that particular visit. The third visit finished with a Preferential Looking Experiment[16] to test comprehension of the eight words taught over the course of all three sessions.

2.3 Vocabulary measures

Parents of all 28 participants filled out the Dutch version of the MacArthur-Bates Words and Sentences Child Development Inventory (N-CDI) at the time of the first recording sessions. For 20 of the 28 participants, an additional follow-up N-CDI was filled out near the child's second birthday. (Comparison to standardized N-CDI word production scores placed five of these 20 in the top quartile and four in the bottom quartile. Thus by age two, our participants appeared representative of the Dutch-learning toddler population.)

Redistribution subject to ASA license or copyright; see http://acousticalsociety.org/content/terms. Download to IP: 195.169.108.229 On: Tue, 26 Nov 2013 11:26:17

2.4 Transcription

The corpus includes over 57 000 utterances, orthographically transcribed using the software program ELAN [http://tla.mpi.nl/tools/tla-tools/elan/; (Ref. 17)]. Utterances were defined as breath groups and were marked as IDS (infant-directed speech), FADS (speech addressed to a familiar adult), UADS (speech addressed to an unfamiliar adult), or RS (read speech). Assignment of utterances to register type was based on reference to both audio and video recordings. Each adult's productions were coded on a separate text tier labeled caregiver 1, caregiver 2, or experimenter. Utterance onset and offset was marked to enable subsequent analyses of speech overlap and turn structure. Instances where transcribers could not identify the words being spoken by the adults were coded as "XX." No attempt was made to transcribe the children's productions.

2.5 Summary data

The corpus contains 29 316 utterances directed to the infant (i.e., IDS), 9834 utterances to a familiar adult (i.e., FADS, or speech between caregivers), 16 754 utterances to an unfamiliar adult (i.e. UADS, or speech between caregiver and experimenter), and 1817 read utterances. The analyses reported in the following text are based on all spontaneous productions made during the entire approximately 30 min recording sessions (excluding the read speech). Caregivers' utterances frequently overlapped even when the infant was being addressed; on average, 27% of the IDS utterances infants heard overlapped with another utterance. Variability in this measure was large, however, with some infants hearing as little as 12% and some hearing as much as 46% overlapping IDS utterances.

Although caregivers did not avoid speaking at the same time to the child, they did clearly modify their speaking style when addressing their child. The effect of speech register (i.e., speaking style used when addressing infants, a familiar adult versus an unfamiliar adult) on utterance length is illustrated in Fig. 1(A). As predicted, utterances were shorter in IDS than in either adult-to-adult speech register. Interestingly, we also found that utterance length differences tended to be greater between UADS and IDS than between FADS and IDS. Speech rate differences between the three registers exhibited a different pattern: Although IDS differed from both UADS and FADS, UADS and FADS speech rates were very similar (see Table 1). Several studies have examined how IDS differs from FADS or UADS, however, this is the first report of which we are aware to treat UADS, FADS, and IDS as three distinct registers. Our overall analyses suggest that studies using only UADS to represent all ADS (as in laboratory studies comparing speech between a mother and child and a mother and experimenter) may lead some IDS/ADS differences to be exaggerated.

Caregivers produced many target tokens [Fig. 1(B)]. Inflected forms were produced more often for adjectives and verbs than for nouns and proper names. Despite the instructions to focus equally on training each word type, caregivers produced fewer tokens of adjectives than nouns, names, or verbs. Given that children's early productions contain few adjectives,[18] caregivers may have strategically focused more energy

Table 1. Average speech rate in syllables per second, by register, caregiver type, and recording session. Only data for the first two sessions are reported because on the third visit, no second familiar adult was present and no grandmothers came in.

	Mother		Father		Grandmother	
	Session 1	Session 2	Session 1	Session 2	Session 1	Session 2
IDS	4.08	4.02	3.89	4.09	3.88	3.87
FADS	4.73	4.82	4.75	5.25	4.87	4.64
UADS	4.70	4.81	4.93	5.32	4.84	4.42

Redistribution subject to ASA license or copyright; see http://acousticalsociety.org/content/terms. Download to IP: 195.169.108.229 On: Tue, 26 Nov 2013 11:26:17

on the word types they thought they had the best chance to successfully teach to their child.

Finally, we examined maternal IDS in relation to individual vocabulary development. Mothers' IDS rates varied greatly (M = 4.1 syllables per second; range = 3.0–5.6). Slower speech rates facilitate young children's word recognition,[14] and we therefore predicted a relation of individual mothers' IDS rate to their children's N-CDI score at 11 months. However, we observed no such relationship of mother's speech rate modification (FADS-IDS difference) with children's measured vocabulary size [see Fig. 1(C)]. Mothers also varied greatly in the proportion of tokens produced in isolation [M = 28%; range = 13% to 46%; note that we included fillers and vocatives in our count, so that this number is higher than in some past studies that excluded such forms (e.g., Ref. 8)]. Prior studies have suggested that exposure to isolated words promotes language development (Ref. 5, the count of which also included all isolated tokens), and we thus predicted that caregivers' production of few isolated words might be linked to smaller productive vocabulary size. However, after excluding the data from one mother who only produced a single token of a trained word, we found the opposite pattern: The production of a greater proportion of multi-word utterances relative to single-word utterances was associated with infants having larger productive vocabularies at 11 months of age, $r(26) = 0.52$, $p = 0.005$. There are at least two plausible explanations for this finding: (1) Mothers modify their speaking style to suit their infants' level of linguistic skill (e.g., Ref. 19) or (2) exposure to multi-word utterances promotes (rather than hinders) vocabulary development in young infants.

3. Conclusion

Children's vocabularies expand at an astonishing rate in the second year of life. This corpus focuses on the language input children are receiving when they are on the brink of this expansion.[18] The corpus is specifically designed to allow researchers to address key questions that are difficult or impossible to address in other corpora examining language input at this crucial stage of development. The findings summarized in this initial report already motivate more nuanced definitions of "adult-directed" versus "infant-directed" speech and provide new insights into the role of early input in word discovery.

Acknowledgments

Corpus construction was funded by NWO-SPINOZA ("Native and nonnative listening"; A. Cutler). The second, third, and fourth authors are affiliated to both the Max Planck Institute of Psycholinguistics and the Radboud University Nijmegen. We thank Tijl Grootswagers for assistance in programming and Angela Khadar for assistance in data acquisition.

References and links

[1]D. R. Mandel, P. W. Jusczyk, and D. B. Pisoni, "Infants' recognition of the sound patterns of their own names," Psychol. Sci. **6**, 314–331 (1995).
[2]E. Bergelson and D. Swingley, "At six to nine months, human infants know the meaning of many common nouns," Proc. Natl. Acad. Sci., U.S.A. **109**, 3253–3258 (2012).
[3]R. Tincoff and P. W. Jusczyk, "Six-month-olds comprehend words that refer to parts of the body," Infancy **17**, 432–444 (2012).
[4]R. Tincoff and P. W. Jusczyk, "Some beginnings of word comprehension in 6-month-olds," Psychol. Sci. **10**, 172–175 (1999).
[5]M. R. Brent and J. M. Siskind, "The role of exposure to isolated words in early vocabulary development," Cognition **8**, B33–44 (2001).
[6]C. Lew-Williams, B. Pelucchi, and J. R. Saffran, "Isolated words enhance statistical language learning in infancy," Dev. Sci. **14**, 1323–1329 (2011).

Redistribution subject to ASA license or copyright; see http://acousticalsociety.org/content/terms. Download to IP: 195.169.108.229 On: Tue, 26 Nov 2013 11:26:17

[7]R. Aslin, J. Woodward, N. LaMendola, and T. Bever, "Models of word segmentation in fluent maternal speech to infants," in *Signal to Syntax: Bootstrapping From Speech to Grammar in Early Acquisition*, edited by J. Morgan and K. Demuth (Erlbaum Associates, Mahwah, NJ, 1996), pp. 117–134.
[8]J. van de Weijer, "Language input for word discovery," unpublished Ph.D. dissertation, Max Planck Series in Psycholinguistics 9 (1998).
[9]D. Burnham, C. Kitamura, and U. Vollmer-Conna, "What's new, pussycat? On talking to babies and animals," Science **296**, 1435 (2002).
[10]A. Fernald, T. Taeschner, J. Dunn, M. Papousek, B. de Boysson-Bardies, and I. Fukui, "A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants," J. Child Lang. **16**, 477–501 (1989).
[11]C. Kitamura and D. Burnham, "Pitch and communicative intent in mother's speech: Adjustments for age and sex in the first year," Infancy **26**, 85–110 (2003).
[12]H. M. Liu, P. K. Kuhl, and F. M. Tsao, "An association between mothers' speech clarity and infants' speech discrimination skills," Dev. Sci. **6**, F1–F10 (2003).
[13]W. Ma, R. M. Golinkoff, D. Houston, and K. Hirsh-Pasek, "Word learning in infant-and adult-directed speech," Lang. Learn. Dev. **7**, 209–225 (2011).
[14]J. Y. Song, K. Demuth, and J. Morgan, "Effects of acoustic properties of infant-directed speech on word recognition," J. Acoust. Soc. Am. **128**, 389–400 (2010).
[15]I. Zink and M. Lejaegere, *N-CDIs: Lijsten voor Communicatieve Ontwikkeling. Aanpassing en hernormering van de MacArthur CDIs van Fenson et al.* (*Communicative Development Inventories. Adaptation and re-standardization of the MacArthur CDI's by Fenson et al.*) (Acco, Leuven, Belgium/Leusden, Netherlands, 2002).
[16]E. K. Johnson and T. Zamuner, "Using infant and toddler testing methods in language acquisition research," in *Experimental Methods in Language Acquisition Research*, edited by E. Blom and S. Unsworth (John Benjamins Publishing, Amsterdam, 2010), pp. 73–94.
[17]P. Wittenburg, H. Brugman, A. Russel, A. Klassmann, and H. Sloetjes, "ELAN: A Professional Framework for Multimodality Research," in *Proceedings of LREC 2006, Fifth International Conference on Language Resources and Evaluation* (2006).
[18]L. Fenson, P. Dale, J. Reznick, E. Bates, D. Thal, and S. Pethick, "Variability in early communicative development," Monogr. Soc. Res. Child Dev. **59**(5), 242 (1994).
[19]E. S. Ko, "Nonlinear development of speaking rate in child-directed speech," Lingua **122**, 841–857 (2012).

Redistribution subject to ASA license or copyright; see http://acousticalsociety.org/content/terms. Download to IP: 195.169.108.229 On: Tue, 26 Nov 2013 11:26:17