Published in final edited form as:

J Cogn Dev. 2005 ; 6(2): 179–208. doi:10.1207/s15327647jcd0602_2.

Dynamics of Word Comprehension in Infancy: Developments in Timing, Accuracy, and Resistance to Acoustic Degradation

Renate Zangl,

Center for Research in Language, University of California, San Diego

Lindsay Klarman,

Center for Research in Language, University of California, San Diego

Donna Thal,

San Diego State University and Center for Research in Language, University of California, San Diego

Anne Fernald, and

Department of Psychology, Stanford University

Elizabeth Bates

Center for Research in Language, University of California, San Diego

Abstract

Online comprehension of naturally spoken and perceptually degraded words was assessed in 95 children ages 12 to 31 months. The time course of word recognition was measured by monitoring eye movements as children looked at pictures while listening to familiar target words presented in unaltered, time-compressed, and low-pass-filtered forms. Success in word recognition varied with age and level of vocabulary development, and with the perceptual integrity of the word. Recognition was best overall for unaltered words, lower for time-compressed words, and significantly lower in low-pass-filtered words. Reaction times were fastest in compressed, followed by unaltered and filtered words. Results showed that children were able to recognize familiar words in challenging conditions and that productive vocabulary size was more sensitive than chronological age as a predictor of children's accuracy and speed in word recognition.

Understanding spoken language seems to be an easy task for adults with robust processing skills, even in difficult listening conditions such as on a bad telephone line or in a crowded room. Adults are efficient listeners not only under normal acoustic circumstances but also in situations where the quality of speech is degraded due to increased speaking rate or reduced spectral information (Lane & Grosjean, 1973; Marslen-Wilson, 1984, 1987, 1993; Speer, Wayland, Kjelgaard, & Wingfield, 1994; Stine, Wingfield, & Myers, 1990). Only when speech is severely distorted does processing get disrupted sufficiently to affect comprehension in the mature listener (Dick et al., 2001; Grosjean, 1985; Munson, 2001; Nooteboom & Doodeman, 1984). The remarkable efficiency of processing in adults has an enormous advantage: The rapid access of word meaning from the acoustic signal prevents conversational breakdowns in normal as well as more difficult listening conditions. Within this context, efficiency can be seen as (a) the ability to rapidly and accurately process incoming speech as well as (b) a higher tolerance to variability in the speech signal. The ability to understand spoken language in more demanding and effortful situations is a critical

capacity in adults; however, little is known about how this robust processing skill develops. In this study, efficiency of processing in the young listener was investigated under normal and potentially adverse processing conditions. To model the latter, we selected two methods, time compression and low-pass filtering, both used previously in research with adults. We hypothesized that (a) these input manipulations would increase the demands of processing and make it harder for the novice listener to understand familiar words and (b) the magnitude of disruption would depend on the child's age and language experience. Because there is no prior work on the effects of these challenging listening conditions on infant online processing, this study was necessarily exploratory in nature.

There are three important reasons for examining the impact of perceptually degraded speech on the young listener. First, children's responses to acoustically modified speech may help researchers to better understand the impact of temporal and spectral variation in the speech signal within a developmental framework. Second, acoustically modified speech is a useful tool for investigating the emergence of robustness in children's spoken-word comprehension. That is, developmental changes in the resilience of word recognition under conditions of acoustic degradation may yield insights into the strength of lexical representations in relation to age and vocabulary growth. Third, experimental measures that reveal processing weaknesses can be adapted for use in clinical contexts to identify children who may have deficits—either in the perception of rapid temporal cues or the perception of certain frequencies due to early hearing loss.

TIME-COMPRESSED SPEECH

Speech varies in its tempo—something all listeners commonly experience. Speakers produce utterances at different rates, ranging on a continuum from relatively slow to relatively fast. Although adults can cope with a high variability in tempo and handle speaking rates about 2 to 2½ times as fast as “normal,” little is known about the impact of variations in speaking rate on the young listener. Although speech addressed to children is usually slower compared to that addressed to adults, children are constantly exposed to variations in speech. For example, adults speak at different tempos and infants overhear a large amount of speech among adults that varies in rate. The design of this study allowed us to directly investigate the impact of variations in tempo by manipulating the speaking rates of familiar words.

In the acoustically unaltered condition, familiar words were presented at their normal rate, with “normal” referring to the speech style commonly known as infant-directed speech (Fernald & Kuhl, 1987; Fernald & Simon, 1984; Fernald et al., 1989). The exaggerated duration and pitch characteristic of infant-directed speech may render these words particularly salient for the young listener.

In the compressed speech condition, the normal speaking rate of the words was reduced by compressing them in time by 50%, making them shorter and closer in duration to words produced in adult-directed speech. Compressing these words by periodically deleting small segments of the signal at regular intervals increased the rate twofold. Because this manipulation preserved crucial spectral information and retained normal rhythmic and prosodic information, the time-compressed words sounded quite natural to the naked ear. Although the intelligibility of the speech signal was relatively unaffected, the young listener was given substantially less time to recognize the words and integrate them into the current context. This increase in processing load could potentially have detrimental effects on word comprehension.

Various studies have examined the effects of manipulating speaking rates on linguistic processing in adults with normal, intact language abilities as well as in those with different types of language disorders (Blumstein, Katz, Goodglass, Shrier, & Dworetzky, 1985;

Foulke, 1971; King & Behnke, 1989; Leonard, Baum, & Pell, 2000; Wingfield, 1996). Fast speaking rates have been shown to have adverse effects on processing, especially at more severe rates of compression, and in older populations in a variety of tasks investigating auditory comprehension, recall, and repetition (Dupoux & Green, 1997; Gordon-Salant & Fitzgibbons, 1999; Schmitt & Carroll, 1985; Wingfield, Tun, Koh, & Rosen, 1999). However, studies examining the effects of variations in speaking rate on linguistic processing in children are scarce and they have primarily focused on children with auditory perceptual language problems, learning disabilities, or acquired aphasia (Campbell & McNeil, 1985; Manning, Johnston, & Beasley, 1977; McNutt & Chi-Yen Li, 1980). The findings of these studies, similar to those for research with adults, suggest that faster speaking rates can negatively affect auditory comprehension and production (Weismer & Hesketh, 1996).

Based on these findings from speaking-rate manipulations, there is reason to believe that slower speaking rates may be beneficial for auditory comprehension. In this study we predicted that if longer duration of the target word facilitated word recognition, infants should do better with normal, acoustically unaltered words than with the compressed variants. If, however, signal duration was not a critical factor in infants' comprehension, this would suggest that infants are capable of handling variations in speaking rate from very early on, which would also be an important finding.

LOW-PASS-FILTERED SPEECH

Speech also varies in spectral composition, which can reduce intelligibility. In certain types of distorted speech the spectral aspects of the acoustic signal are affected. The reduction of spectral information in the speech signal disrupts the encoding of the linguistic information, making the signal less intelligible and comprehension more effortful. Adults show perceptual resilience in the face of many forms of spectral distortion (Aydelott & Bates, 2004; Remez, Rubin, Pisoni, & Carrell, 1981; Shannon, Zeng, & Wygonski, 1998; Shannon, Zeng, Wygonski, Kamath, & Ekelid, 1995; Turner, Souza, & Forget, 1995). The question is how this resilience develops and what potential advantages it has. An important result of gaining this skill is that even with less redundancy in the speech signal the listener can still pick up enough acoustic information to succeed in comprehension. The relevant question here is how early children develop the ability to understand speech reduced in its spectral content and whether this ability develops gradually over the age range of this study. That is, as children get older and more linguistically sophisticated, do they depend less on the integrity of spectral cues to correctly identify words?

By using a low-pass filter we attenuated all frequency components above 1,500 Hz. Although the prosodic and temporal aspects of the words stayed intact, the words sounded muffled to the adult ear and phonemic contrasts became less clear. Even though the low-pass-filtered words lacked several features that may contribute to the identification of vowels and consonants, there was still sufficient information for experienced listeners to quickly and accurately identify the words (as we ascertained in a pilot study with 55 college students).

Previous studies have shown that less language-experienced listeners are particularly vulnerable to the reduction of spectral information in the speech signal. Testing two groups of children (5–7 years and 10–12 years) as well as adults in a variety of tasks (sentence–word comprehension, recall of digits, etc.), Eisenberg, Shannon, Schaefer Martinez, and Wygonski (2000) demonstrated that the younger children performed at a significantly lower level in all tasks compared to the older children and adults. These results are in accordance with an earlier study indicating that children age 3 to 4 years need more spectral information

than adults do to understand multisyllabic words (Dorman, Loizou, Kirk, & Svirsky, 1998). The fact that younger children are more vulnerable to spectrally degraded speech suggests that robust speech recognition requires substantial experience.

If more spectral information in the speech signal is critical for early word recognition, then the youngest listeners in our study should do better with normal, acoustically unaltered words than with the spectrally impoverished variants. However, if we find that reduction of spectral information above 1,500 Hz does not diminish comprehension, this would indicate that language novices are capable of handling degraded spectral content from very early on. This would be an important finding, especially in relation to children who suffer from early hearing loss and receive imperfect sensory input.

EFFICIENCY, AGE, AND VOCABULARY

We approached the question of efficiency in word recognition and resistance to challenges imposed by time compression and low-pass filtering from two points of view: comparisons over levels of age and comparisons over levels of expressive vocabulary. To the extent that increasing efficiency of word comprehension reflects maturational factors, we might expect to find effects of age that are independent of vocabulary level. However, increasing efficiency in receptive language skills may also be related to gains in productive vocabulary. Previous studies using a range of methodologies have shown that several aspects of language (especially word comprehension and grammar) are correlated with the infant's level of expressive vocabulary after age-related variance is removed (Bates & Goodman, 1997; Fernald, Swingley, & Pinto, 2001; Marchman & Bates, 1994; Mills, Coffey-Corina, & Neville, 1993, 1997; Munson, in press; Werker, Fennell, Corcoran, & Stager, 2002). Whereas these studies indicate that both age and vocabulary contribute to the increase in infants' processing efficiency, one of our goals was to establish whether word comprehension would be predicted better by expressive vocabulary level or by chronological age.

Recent research investigating spoken-word comprehension has suggested that efficiency in processing develops gradually and depends on the child's age and vocabulary knowledge as well as on the listening conditions. Age-related changes have been reported by Fernald, Pinto, Swingley, Weinberg, and McRoberts (1998) showing that 24-month-olds were both more accurate and faster than 15-month-olds in recognizing familiar spoken words (for related studies see also Mills et al., 2004; Werker & Stager, 2000; Werker et al., 2002). In terms of the relation between word comprehension and vocabulary size, links between receptive and productive skills have been documented through observational studies and more recently by experimental research. In a longitudinal study of the emergence of speech-processing efficiency in children age 15 to 25 months, Fernald, Perfors, & Marchman (in press) found that speed and accuracy in spoken language comprehension were closely related to vocabulary development. At any given age, those children with larger vocabulary size were faster and more reliable in recognizing familiar words.

In another recent study, Werker et al. (2002) found effects of both age and vocabulary size when investigating minimal-pair word learning in infants at 14, 17, and 20 months of age. When examining the relation between vocabulary size and word learning, they found that those children in each age group with larger vocabularies were better in learning similar-sounding words. However, a positive correlation between minimal word learning and vocabulary size was found only in the younger infants with smaller vocabularies (see also Werker & Stager, 2000). Although Werker et al. investigated learning of new words rather than recognition of already familiar words, the results bear important implications for this study. They showed that children became more efficient in processing phonetic details and

that vocabulary size may be important in the development of efficiency (for further studies relevant for the contribution of lexical growth, see Fernald et al., 2001). Taken together, these studies provide preliminary evidence that efficiency in receptive processing may be related to gains in productive vocabulary.

In this study, the relation between receptive processing skills and productive vocabulary size was extended to a broader range of speech sounds, including unaltered, time-compressed, and low-pass-filtered words. To explore the relative contributions of increasing age and growth in productive vocabulary size to the development of speed and accuracy in spoken-word comprehension, we grouped all our analyses both by age and by vocabulary level. To determine how success in online word recognition varied with age, the children were divided into four age groups. To determine how their word recognition success varied with level of lexical development, the same children were grouped into four vocabulary levels. The size of the expressive vocabulary for each child was based on the number of spoken words reported by the parent on the MacArthur-Bates Communicative Developmental Inventory (CDI).

In addition to effects of age and vocabulary size, efficiency in speech processing is also influenced by the listening conditions to which the young learner is exposed. Recent research on the online processing of spoken language by infants in the 2nd year of life has shown that young children can efficiently recognize familiar words not only under optimal (i.e., acoustically unaltered) conditions but also under more challenging conditions such as when familiar words are truncated or mispronounced. For example, 18-month-olds are able to identify familiar words based on partial phonetic information, when presented with only the first 300 msec of the word (Fernald et al., 2001). And infants also recognize words that deviate slightly from the correctly pronounced version, as in *vaby* versus *baby* (Mills et al., 2004; Swingley & Aslin, 2000, 2002). These findings indicate that language novices are able to handle instances of the same word in different acoustic manifestations—a fact particularly relevant in this research.

GOALS OF THIS STUDY

In a population of 95 English-learning children ranging in age from 12 to 31 months, we investigated the development of efficiency in comprehending familiar words under optimal and less optimal conditions. Children were tested in a looking-while-listening procedure, a technique that has been used in a number of recent studies on infant word recognition (Fernald et al., 1998, 2001; Swingley & Aslin, 2000, 2002). This procedure allowed us to monitor the time course of spoken-language processing as the speech signal unfolded from moment to moment. The child looked at paired pictures of familiar objects while listening to speech that referred to one of the objects. By examining the child's eye movements continuously in response to particular words in the speech stream, we were able to tap into the rapid mental processes involved in understanding spoken language. The advantage of this procedure is that language processing can be monitored as it happens—while the child is listening to speech in real time. Previous studies of early receptive language skills have relied primarily on offline measures that assess word recognition only *after* the child has heard the entire sentence. In contrast, the looking-while-listening procedure enables measurement of receptive competence in terms of response speed as well as overall accuracy in spoken word recognition.

This study had three goals:

1. To investigate the development of efficiency in comprehending familiar words under optimal listening conditions. This allowed us to replicate the earlier findings of Fernald et al. (1998) indicating that the efficiency in processing of acoustically

normal words increases gradually over the 2nd year of life. Our study extended the age to a wider range, testing children age 12 to 31 months.

2. To investigate the development of efficiency in comprehending familiar nouns when they were presented either at a faster speech rate or with reduced spectral information. This would clarify whether manipulations in rate and clarity affect early word comprehension and would also provide insights into the robustness of word comprehension in young language learners.
3. To determine whether developmental changes in performance within the first stages of word learning are best predicted by age or vocabulary level or by these two factors in conjunction.

METHOD

Participants

Children were recruited through bulk mailing, brochures, advertisements in local parent magazines, and visits to postnatal classes. All of the infants who came to the laboratory were full-term and in good health, with neither pre- nor postnatal complications nor a history of hearing disorders. Our final sample consisted of 95 children age 12 to 31 months, 54 girls and 41 boys, all from monolingual English-speaking homes. A breakdown of the number of participants within each age level is provided in Table 1.

An additional 24 infants were tested but not included in the analyses for the following reasons: failure to complete the task ($n = 11$), missing parent-report language inventories ($n = 6$), experimenter error or equipment failure ($n = 3$), difficulty in tracking the child's eye movements ($n = 1$), interference by parents during testing ($n = 1$), fussiness ($n = 1$), and failure to meet the testing criteria ($n = 1$).

Auditory Stimuli

Twenty-four target words were chosen from the earliest words comprehended by typically developing children, based on the norms of the MacArthur-Bates CDI (Fenson et al., 1993). Table 2 lists all the target words used in the experiment. The auditory stimuli were digitally recorded in a soundproof room by a female native speaker of American English, at a sampling rate of 44,000 Hz using a Sony digital audiotape recorder. The acoustic envelope of each word was typical of infant-directed speech, showing both extended duration and pitch patterns (e.g., Fernald et al., 1989). The speech stimuli were then digitized at 22,050 Hz using Sound Designer (Digidesign) for Macintosh and converted into 16-bit wav. files for use on a Windows-DOS system. To direct the child's attention to the target picture, a standard, invariant carrier frame preceded each target word ("Look, look at the [target]"). To create the carrier phrase, the initial "Look" was spliced into each sentence; the rest of the sentence was recorded separately for each target word in a form designed to maximize the naturalness of the lead-in phrase while minimizing coarticulation effects. Whereas the carrier frame was always presented in normal, acoustically nonmodified speech, the target words had three acoustic definitions: (a) unaltered (normal), (b) low-pass-filtered, and (c) time-compressed. The acoustic distortions were imposed on the target words by using the Equalizer function for low-pass filtering and the Tempo function for time compression in Sound Edit 16 (Macromedia). Equalizer changed the spectral resolution of the speech signal by eliminating all frequency information above 1,500 Hz. Tempo decreased the original stimulus length by 50% but preserved segmental and pitch information. The level of degradation for the acoustically altered target words used in this experiment was based on a prior study with adults.¹ The mean length of the unaltered and filtered stimuli was 1,051.85

msec (range: 820–1,409 msec); the mean length of the compressed stimuli was 520.58 msec (range: 390–721 msec).

Although both filtered and time-compressed tokens were prepared for all of the stimulus words, each child heard only one of the degraded versions of each word in addition to the unaltered version. That is, each target word was presented in two forms to each child—once unaltered and once either time-compressed or low-pass-filtered and each child heard equal numbers of filtered and time-compressed stimuli. The total of 48 trials was split up into eight blocks of 6 trials with each block containing three unaltered and three acoustically modified targets, with the type of distortion constant within but variable across blocks. Thus each child was exposed to a total of 24 unaltered and 24 perceptually degraded targets (12 time-compressed and 12 low-pass-filtered). The pairings of targets and distracters were based on phonological and semantic dissimilarity with the criterion that within a given pair the words could neither come from the same semantic category nor start with the same initial phoneme. Furthermore, targets and distracters were also matched in relation to the age at which they were first typically comprehended and produced based on MacArthur-Bates CDI norms (e.g., an easy target had an easy distracter and vice versa, as in *dog* and *car* vs. *pig* and *hat*).

Visual Stimuli

The visual stimuli were 16-bit digitized realistic images of early-learned objects in 300 × 200 pixel size, presented side by side on two 30-cm color video monitors. The object matching each spoken target word had four visual exemplars, all prototypical instantiations of the respective word approximately balanced for visual salience. The images were downloaded from CD ROMs or the Internet or derived from scanned digital photographs and edited on Adobe Photoshop. Each picture served twice as a target and twice as a distracter, resulting in a total of 96 pictures. Target and distracter pictures appeared simultaneously on the screens and were presented 650 msec before the onset of the sentence and a total of 2,050 msec before the onset of the target word. The pictures stayed on through the entire auditory event and beyond; picture off-set was at 5,250 msec.

Apparatus and Procedure

Each child was tested in a soundproof room. The procedure used for data collection was the standard preferential looking procedure described by Schafer and Plunkett (1998). During testing, the child was seated centrally on the parent's lap, 80 cm in front of a pair of 30-cm computers placed about 44 cm apart from each other. Speech stimuli were delivered at around 70 dB through a concealed speaker located centrally above the monitors. Children's looking behavior was recorded by two cameras, one located above the right monitor, the other one located above the left monitor. Video feed from both cameras was recorded onto two VHS videotapes by using a split-screen option on an audiovisual mixer.

Before testing, the parent was given an introduction to the purpose and nature of the study by one experimenter while the other experimenter entertained the child. Additionally, the parent signed a consent form and submitted the MacArthur-Bates CDI, which had been filled out at home (MacArthur-Bates CDI: Words and Gestures from 12–16 months; Words and Sentences from 17–31 months; Fenson et al., 1993). When the child was at ease, they were led into the testing room with their parent. The parent was seated in a chair, wearing opaque dark glasses so she could not see the target pictures and listening to masking music

¹A prior study with 55 college students was carried out using the same target words as in the experiment presented here but two levels of degradation (1,000 Hz and 1,500 Hz for low-pass filtering and 50% and 100% for time compression). Adults reliably recognized the target words (with mean accuracy scores above 95%) in time compression as well as low-pass filtering in either degradation level.

over headphones throughout the experiment. These procedures were designed to prevent the parent from cueing the child regarding the location of the named picture on each trial. A small red light and a buzzer were mounted between the monitors. These served as “attention-getting” devices between trials to redirect the child’s attention to the center and away from both monitors. Each trial started with the presentation of two pictures, so the child had enough time to look at both pictures prior to the onset of the target word despite the use of the central fixation light. When the target word was presented, the child could be in one of three positions: already looking at the correct picture, looking at the incorrect picture, or off-task. The experimenter in the adjacent room advanced from one trial to the next only after determining that the child was looking at the light. The experiment lasted 8 to 10 min on average.

Coding Eye Movements

Although the data were collected using the Schafer and Plunkett (1998) procedure, we used a combination of methods for coding and analyzing the data. Following Schafer and Plunkett, each recording session was scored offline using a button-press box. Two highly trained coders independently scored each session two times, once for looks to the right monitor and once for looks to the left monitor. The coders were unaware of the trial type and the position of the target and distracter pictures on each trial. Reliability was assessed both within and between observers following procedures described by Schafer and Plunkett. Across the 95 children in this study the mean percent agreement was 95% or higher in all sessions.

Because we used a button box to code infants’ gaze patterns to the two pictures, it is important to note that the response latency of the coder was included in the measurement of looking time, as is the case in several well-established versions of the preferential looking procedure (e.g. Golinkoff, Hirsh-Pasek, Cauley, & Gordon, 1987; Hollich et al., 2000; Naigles, 1990; Schafer & Plunkett, 1998). In such studies the preferred dependent measure is typically the proportion of looking time to the named target picture over a circa 6-sec window following the offset of the target word. However, because we were interested in getting more detailed information on the time course of infants’ looking time, we chose to use a different approach for analyzing our data, following procedures developed by Fernald and Swingley (e.g., Fernald et al., 1998; Swingley, Pinto, & Fernald, 1999). As in research on spoken-language processing with adults (e.g. Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995), these investigators monitored infants’ looking time from moment to moment, combining “window” analyses that averaged looking time over a particular period of interest with reaction time (RT) analyses that captured the speed of the child’s online recognition of the target word.

Although our data were coded offline in real time using the button box rather than frame-by-frame (see Fernald et al., 2001; Swingley&Fernald, 2002), we subsequently divided the looking-time data into 25-msec windows using custom software for purposes of further analysis. This enabled us to achieve a more fine-grained record of children’s gaze patterns in response to the spoken target words. For each 25-msec window it was noted whether the child was fixating the target or distracter picture, was in a transition shifting from one picture to the other, or was off-task (looking away). These measurements were then analyzed in relation to the onset of the target word on each trial.

Dependent Measures

Children’s efficiency in recognizing spoken target words under different perceptual conditions was assessed using measures of accuracy in word recognition and latency to orient to the matching picture in response to the spoken target word.

Accuracy—To determine the accuracy of children's fixations to the correct picture in response to target words in unaltered, time-compressed, and low-pass-filtered form, we measured their looking behavior over time as more and more phonetic information specifying the target word became available. Following Swingle, Pinto, and Fernald (1998), we analyzed looking times using two 1-sec time windows starting from the onset of the target word, with *accuracy* defined as the time children spent looking at the target picture as a proportion of the total time spent looking at both target and distracter pictures. If infants' tendency to look at the two pictures was influenced by hearing the target word, we would expect a substantial increase in looking to the target picture in the second 1-sec window as compared to the first 1-sec window, as more phonetic information unfolded. In a study of word recognition by infants age 18 to 21 months, Fernald et al. (2001) showed that 82% of correct shifts from the distracter to the target picture occurred within the first 1,800 msec following target word onset, supporting the prediction that looking time to the target picture would reach its peak in the second 1-sec time window in our data as well.

RT—Infants' speed in recognizing familiar spoken words was measured in terms of their average latency to initiate a shift from the distracter to the target picture in response to the target word. RT calculations were based on those trials on which the child shifted correctly from the distracter to the target picture. Following Fernald et al. (2001), we divided our trials into three categories depending on where the child was looking at target word onset: T-initial trials (when children were already looking at the target picture at word onset), D-initial trials (when children were looking at the distracter picture at target word onset), and A-trials (when children were looking away from either picture). Across the whole data set, 45% of all trials were classified as D-initial trials, 41% were T-initial trials, and 14% were A-trials. A-trials were excluded from all further analyses. For the D-initial trials, the *time window* during which shifts from the distracter to the target picture were considered to be correct was defined as extending from 625 msec to 2,000 msec. The lower cutoff of 625 msec accommodated both the time required to disengage from one picture and initiate a shift to another picture (ca. 300–360 msec according to Haith, Wentworth, & Canfield, 1993) as well as the scorer's estimated latency to press the button on the button box (ca. 300 msec). The upper cutoff of 2,000 msec used in these analyses was consistent with previous research by Swingle et al. (1999).

Grouping Variables

Children's age and level of lexical development were the two grouping variables of interest in this research. By grouping the same children first by age and then by vocabulary size, our goal was to determine the extent to which either or both of these factors was associated with the accuracy and speed with which infants were able to identify spoken words under different conditions of stimulus integrity. For the analyses by age, children were grouped into four age levels: 12 to 14 months ($n = 22$, M age = 12.6), 15 to 18 months ($n = 25$, M age = 16.4), 19 to 23 months ($n = 30$, M age = 20.6), and 24 to 31 months ($n = 18$, M age = 25.7). For the analyses by vocabulary, the same children were reassigned to another set of four groups: 0 to 20 words ($n = 27$, M words = 6.9), 21 to 99 words ($n = 31$, M words = 52.9), 100 to 300 words ($n = 21$, M words = 183.2), and more than 300 words ($n = 16$, M words = 462.4). These particular age and vocabulary groupings were chosen to reflect windows of maximal homogeneity within groups and maximal change between groups, based on our previous behavioral work on vocabulary development across this age range (Bates & Goodman, 1997; Caselli, Bates, et al., 1995; Caselli, Casadio, & Bates, 1999). The split into four vocabulary groups was based on previous research using the MacArthur-Bates CDI as well as careful preliminary analyses of the MacArthur-Bates CDI data for our sample. Both suggested breakpoints in the developmental curves for vocabulary production at the following levels: more than 20 words, more than 100 words, and more than 300

words. Within the 0-to-20-word level, words are typically integrated in a piece-meal fashion as the child's vocabulary expands at a slow rate. The rate of growth catches up at the next level with faster integration of new words typically from around 50 words onward. Between 100 and 300 words, word combinations get off the ground along with more rapid integration of new words. Above 300 words the development of grammar is well underway; the children's language has increased not only in mean length of utterance but also in the use of grammatical morphemes. In the same vein, these age groupings reflect the average windows within which changes in the rate of learning as well as the composition of vocabulary have been observed in numerous previous studies (e.g., Caselli, Casadio, & Bates, 2001).

RESULTS

Accuracy of Word Recognition

For each child mean accuracy scores were calculated for responses in each of the three perceptual conditions, based on the proportion of looking time to the target picture in each of the two 1-sec time windows. The goal of the first analysis was to verify that children's looking to the target picture increased from the first to the second 1-sec time window, indicating that they oriented selectively to the matching picture as the target word unfolded. To confirm that there was indeed a difference between the first and second time windows, we conducted a 2 (time window) $\times$ 3 (perceptual condition) within-subjects analysis of variance (ANOVA). This analysis yielded significant main effects of time window, $F(2, 188) = 107.25, p < .0001$, and perceptual condition, $F(2, 188) = 24.44, p < .0001$, as well as a Time Window $\times$ Perceptual Condition interaction, $F(4, 376) = 14.02, p < .0001$. The main effect of time window confirmed that the mean proportion of looking to the target was significantly greater in the second time window ($M = .60, SE = .01$) than in the first ($M = .48, SE = .007$), demonstrating that target recognition increased over time.

The main effect of perceptual condition revealed that children's accuracy varied with the acoustic integrity of the target word. Accuracy scores were highest in response to unaltered speech ($M = .58, SE = .009$), somewhat lower with compressed speech ($M = .55, SE = .01$), and considerably lower with low-pass-filtered speech ($M = .49, SE = .009$). Follow-up tests showed that the difference in accuracy scores was significant between unaltered and compressed, unaltered and low-pass-filtered, as well as compressed and low-pass-filtered words, $ps < .05$. The Time Window $\times$ Perceptual Condition interaction is illustrated in Figure 1. To explore the source of the interaction, we conducted separate one-way ANOVAs for each time window with perceptual condition as a within-subjects variable. The analyses yielded a main effect of perceptual condition in the first 1-sec time window, $F(2, 188) = 5.70, p < .005$, as well as in the second 1-sec time window, $F(2, 188) = 30.49, p < .0001$. Follow-up tests were conducted to further examine the main effect of perceptual condition in each time window. In the first 1-sec time window, accuracy scores were significantly higher in unaltered compared to compressed and low-pass-filtered words, $ps < .05$. In the second 1-sec time window, accuracy scores were significantly higher in unaltered and compressed compared to low-pass-filtered words, $ps < .05$. The difference between unaltered and compressed words was no longer significant. Because children's accuracy in target recognition peaked in the second 1-sec time window, we focused exclusively on this time window in all subsequent analyses.

The goal of the next set of analyses was to examine children's accuracy in responding to the target word first in relation to age and then in relation to vocabulary size. Starting with the analysis by age, we compared the mean proportions of correct looking time to the target picture during the second 1-sec time window in a 4 (age group) $\times$ 3 (perceptual condition) mixed ANOVA, with age group as a between-subject variable and perceptual condition as a within-subjects variable.

The main effect of age group was significant, $F(3, 91) = 10.24, p < .0001$, reflecting an increase in performance with increasing age, from a mean of .52 ($SE = .018$) at 12 to 14 months to a mean of .66 ($SE = .02$) at 24-plus months. There was also a significant main effect of perceptual condition, $F(2, 182) = 36.44, p < .0001$. Accuracy was highest overall in the unaltered speech condition ($M = .65, SE = .01$), slightly lower in the compressed speech condition ($M = .63, SE = .018$), and worst in the low-pass-filtered speech condition ($M = .52, SE = .014$). Follow-up tests showed that the accuracy scores differed significantly between unaltered and low-pass-filtered, as well as between compressed and low-pass-filtered words, $ps < .05$. The interaction of Age Group $\times$ Perceptual Condition was also significant, $F(6, 182) = 4.30, p < .0004$. As illustrated in Figure 2, this interaction reflects the finding that the older children responded more accurately than the younger children to both unaltered and compressed speech, whereas performance hovered near chance across all ages for filtered words. Post hoc Tukey honestly significant difference (HSD) test analyses, $ps < .05$, showed significant improvements in accuracy between the 12-to-14-month-old and 15-to-18-month-old age groups, and between the 15-to-18-month-old and 24-to-31-month-old age groups, on unaltered speech trials. Accuracy on the compressed word trials also increased reliably from the 12-to-14-month-old age group to the 15-to-18-month-old group and then continued to increase gradually, although improvements were not significant. In contrast to children's responses to unaltered and compressed target words, their success in recognizing target words in low-pass-filtered speech did not change at all across age levels.

In a parallel analysis on the same data, the children were grouped by vocabulary size instead of age. As illustrated in Figure 3, this analysis yielded comparable results with significant main effects of vocabulary level, $F(3, 91) = 13.06, p < .0001$, and perceptual condition, $F(2, 182) = 35.41, p < .0001$, as well a significant Perceptual Condition $\times$ Vocabulary Level interaction, $F(6, 182) = 4.12, p < .001$. Although the direction and the magnitude of the interaction were similar to those of the interaction that emerged when children were grouped by age, there was an important difference. Contrary to our findings when children were grouped by age level, grouping the children by vocabulary size revealed significant differences in target recognition in all three conditions, including the low-pass-filtered condition: unaltered, $F(3, 91) = 14.36, p < .0001$; compressed, $F(3, 91) = 7.36, p < .0001$; low-pass-filtered, $F(3, 91) = 4.03, p < .01$. In fact, post hoc Tukey HSD test analyses ($ps < .05$) indicated that children in the 100-to-300-word group were significantly better at recognizing low-pass-filtered words than were children in the 21-to-99-word group. Unexpectedly, we found that children's success in recognizing low-pass-filtered words dropped for the highest group (i.e., those with more than 300 words), a discrepancy to be considered further in the Discussion.

In summary, these analyses of children's accuracy in online recognition of spoken target words showed that success in word recognition varied with the acoustic integrity of the speech stimulus and with both the age and vocabulary size of the child. Taken together, accuracy in target recognition was higher in unaltered and compressed words as compared to low-pass-filtered words. Target recognition in unaltered and compressed words improved both with age and expressive vocabulary size. Increased performance in low-pass-filtered words was evident only when children were grouped by vocabulary size, and then only in children in the 100-to-300-word group.

In the previous analyses we found that grouping children by vocabulary size revealed group differences that did not emerge when the same children were grouped by age. This finding suggests that vocabulary size may be more sensitive than age in predicting differences in children's success in identifying words that vary in acoustic integrity. However, age and vocabulary are of course highly correlated in our sample of children ($+0.71, p < .0001$). As shown in the cross-tabulation of age and vocabulary groupings in Table 3, most of the

children fall on the diagonal. Thus it would not be legitimate to use both age and vocabulary size as factors in the same design. The only way to evaluate the relative contributions of age and vocabulary size is in a multivariate design that examines the effect of one variable while controlling for the other. Thus we conducted a final set of analyses on the accuracy data to determine whether age and vocabulary levels accounted for the same or separable variance in our looking-time measures. Each of the ANOVAs reported earlier was repeated as an analysis of covariance (ANCOVA). In the age group analysis, vocabulary level was used as a covariate; in the vocabulary group analysis, age group was used as a covariate.

In the ANCOVA with age group as the between-subject factor, the effect of age on accuracy was no longer significant when vocabulary level was controlled, $F(3, 90) = 1.85, ns$. However, a significant contribution from the vocabulary covariate was noted, $F(1, 90) = 11.41, p < .001$. The significant effect of perceptual condition and the Age Group $\times$ Perceptual Condition interaction found in the original analysis were confirmed in the ANCOVA. Because the vocabulary level covariate did not interact with perceptual condition, we can be confident that assumptions on the use of covariance analysis were respected here. We can thus conclude that children's vocabulary size was a more important factor than age in determining their overall accuracy in spoken-word recognition.

In the converse analysis using vocabulary level as the between-subject factor and age group as a covariate, the main effect of vocabulary level on accuracy continued to reach significance, $F(3, 90) = 3.91, p < .01$, although the contribution of the age group covariate as a main effect did not reach significance, $p < .13$. The fact that age group did not interact significantly with perceptual condition again indicates that we have respected the assumptions of analysis of covariance. However, this ANCOVA produced somewhat different results from the original vocabulary level analysis without a covariate. Whereas the main effects of perceptual condition and the Perceptual Condition $\times$ Vocabulary Level interaction did reach significance in the original analysis, both failed to reach significance when continuous effects of age were controlled: perceptual condition, $F(2, 180), p < 1.0, ns$; Perceptual Condition $\times$ Vocabulary Level, $F(3, 180), p < 1.2, ns$. We can thus conclude that the main effect of vocabulary level was more robust than the effect of age group in our analyses of children's accuracy in target word recognition, although both factors were influential.

Speed of Word Recognition

Our next goal was to examine the effects of age, vocabulary size, and stimulus integrity on speed of word recognition by children age 12 to 24-plus months. We thus conducted a series of analyses on RT that paralleled the analyses of accuracy scores presented in the previous section. Having established that accuracy in word recognition varied with the integrity of the target word as well as with age and vocabulary size, our first question was whether infants' response times also differed as a function of perceptual condition. It should be noted that mean RT data was not available for all participants in all three perceptual conditions. Twenty-two of the 95 children in the sample had missing means in one or more of the three conditions and thus could not be included in the analyses. Again, two sets of analyses were conducted—one in relation to age and one in relation to vocabulary size.

In the first analysis we examined the mean RT of children's responses to spoken target words as a function of stimulus integrity and age group. A 4 (age group) $\times$ 3 (perceptual condition) mixed ANOVA revealed a significant main effect of perceptual condition, $F(2, 138) = 6.80, p < .01$. Examination of the cell means showed that RTs were significantly faster in the compressed speech condition ($M = 1,062$ ms; $SE = 33$) as compared to both the unaltered ($M = 1,144$ msec; $SE = 27$) and low-pass-filtered ($M = 1,192$ msec; $SE = 32$) conditions at $ps < .05$. RTs to normal and filtered words did not reliably differ from each

other. The finding that children responded fastest to compressed words is interesting because in this condition complete phonetic information specifying the target word was available earlier than in unaltered speech. However, contrary to our predictions, the main effect of age group failed to reach significance, $F(3, 69) = 2.09$, *ns*. Thus, in contrast to our finding of age-related differences in accuracy scores, we did not find differences in RT using chronological age as a grouping variable. The Age Group $\times$ Perceptual Condition interaction was not significant.

Next we conducted the parallel analysis of mean RTs to spoken target words with vocabulary level as the between-subject factor. Whereas the effect of age group was not significant in the previous analysis, the main effect of vocabulary level on differences in RT was significant, $F(3, 69) = 5.21$, $p < .01$, reflecting a decrease in mean RT across the four vocabulary levels: fewer than 20 words ($M = 1,207$ msec; $SE = 43$), 21 to 99 words ($M = 1,196$ msec; $SE = 26$), 100 to 300 words ($M = 1,082$ msec; $SE = 34$), and more than 300 words ($M = 990$ msec, $SE = 40$). Post hoc tests showed that the difference in mean RT between the lowest and the highest vocabulary levels was significant, with no significant differences between the other groups.

When children were grouped by vocabulary level, the main effect of perceptual condition was also significant, $F(2, 138) = 8.20$, $p < .0004$, just as we found when children were grouped by age. To examine further the effects of stimulus integrity on RT, we conducted follow-up tests within each perceptual condition. One-way ANOVAs were conducted for each of the three conditions, first grouping the children by age and then by vocabulary level. We found significant effects for age group, $F(3, 69) = 5.82$, $p < .001$, and vocabulary level, $F(3, 69) = 7.05$, $p < .0003$, but only in children's responses to unaltered speech. Post hoc tests showed that children significantly increased their response speed from the 15- to 18-month (or 20–99-word) group to the 24-plus-month (or 300-plus-word) group. The gradual decrease in RT from 15 to 18 months onward is consistent with earlier findings by Fernald et al. (1998) in children ranging from 15 to 24 months of age. Note that the *absolute* differences in mean RTs between our results and those reported by Fernald et al. (1998) can be explained by the fact that we used a button-press box in this study rather than coding frame-by-frame, thus adding the circa 300-msec response latency of the observer to children's "raw" response times.

Although the effects of age and vocabulary level were significant only for unaltered speech, there were marginally significant effects of vocabulary level for both compressed and low-pass-filtered speech, $ps < .08$. For compressed words, mean RTs dropped from 1,162 msec in the 21- to 99-word group to 945 msec in the 300-plus-word group. For filtered words as well, the effect of vocabulary level was marginally significant, reflecting a gradual decrease in RT with increasing vocabulary. Mean RTs dropped from 1,331 msec in children with fewer than 20 words to 1,087 msec in children with more than 300 words. The important point here is that vocabulary grouping had a reliable overall effect on RTs whereas age did not, although the most straightforward patterns of change related to either vocabulary size or age were found for unaltered speech.

These analyses suggest that for RT differences just as for differences in accuracy, vocabulary size might be a stronger predictor of children's performance than is age. To test this hypothesis more rigorously, we once again used separate ANCOVAs in which children were grouped either by age or by vocabulary size, with the other grouping variable included as a covariate, parallel to our ANCOVA analyses of accuracy scores. In the ANCOVA using age group as the between-subject variable, there was no main effect of age group after vocabulary level was controlled, $F(3, 68)$, $p < .13$, *ns*. However, there was a significant contribution of vocabulary level as a covariate, $F(1, 68) = 7.22$, $p < .009$. No other main

effects or interactions reached significance in this analysis, suggesting that controlling for vocabulary level removes detectable effects of stimulus integrity and age on RTs. In the corresponding ANCOVA using vocabulary level as the between-subject variable, a main effect of vocabulary level on RTs remained despite the control for levels of age group, $F(3, 68) = 2.97, p < .04$, whereas age group failed to make a significant contribution as a covariate, $p < .90, ns$. Again, no other significant main effects or interactions were observed. An examination of the marginal means for vocabulary level in this covariance analysis (i.e., the estimated means after levels of age group were controlled) revealed a pattern similar to the one reported for the simple ANOVA: mean RT for children with 20 or fewer words was 1,204 msec ($SE = 53$); mean RT for children with 21 to 99 words was 1,196 msec ($SE = 37$); mean RT for children with 100 to 300 words was 1,083 msec ($SE = 45$); and mean RT for children with more than 300 words was 994 msec ($SE = 60$). These findings indicate that children's speed of response in spoken-word recognition is more strongly related to productive vocabulary size than to age.

DISCUSSION

Studies of speech processing under adverse conditions by adults and older children have shown that more taxing processing conditions affecting the integrity of the speech signal—such as temporal acceleration, low-pass filtering, and addition of white noise—can put speech comprehension at risk. As a result, responses may get less accurate and slower. However, the ability of the young, language-learning child to recognize speech under less optimal processing conditions is not well understood. This study addressed this question by testing infant word recognition under optimal conditions, in which the full acoustic spectrum was provided, and less optimal conditions, in which speech was either time-compressed or low-pass-filtered.

One major finding of this research was that infants could recognize words even under suboptimal conditions. They were able to recognize time-compressed as well as low-pass-filtered words. However, the efficiency of word recognition under these challenging conditions was dependent on the type of perceptual degradation as well as on the child's age and expressive vocabulary size. Word recognition was best when the speech signal was unmodified, slightly lower when the speech signal was accelerated, and severely disrupted when the speech signal was poor in its spectral information.

Accuracy in Word Recognition

With regard to accuracy, time compression had a less severe impact on children's ability to recognize the labeled object than did low-pass filtering. Whereas the recognition of time-compressed words was surprisingly efficient from relatively early on, the recognition of low-pass-filtered words was quite fragile even in children with higher expressive vocabularies. Low-pass filtering a word that was identified without problems in its unaltered form disrupted the normal response to the word to a degree unexpected in its magnitude. This difference was much more drastic than in a previous pilot study of adults (even when using more severe degradation levels in each condition). Studies of older children by Eisenberg et al. (2000) and Dorman et al. (1998), who observed long-term developmental learning effects under conditions of spectral degradation, have suggested that reduced spectral information poses severe problems for the young learner. In view of their results it is not surprising that low-pass filtering had an even more deleterious effect on infants. It remains to be determined how much degradation is required to set off these effects. Using word-initial mispronunciations, Swingley and Aslin (2000) documented significant reductions in speed and accuracy when recognizing target words but the effects observed in their study were much milder than those observed in ours. The mild costs of time compression compared to the severe costs of low-pass filtering suggest that signal length

and signal clarity play different roles in early word comprehension, indicating that young word learners are much more vulnerable to poor acoustic quality of the speech signal than to variations in its length, at least within the ranges tested here.

Speed of Word Recognition

With regard to RTs, several findings emerged. First, somewhat surprisingly, the fastest RTs were observed in children's responses to compressed rather than unaltered speech. However, as predicted, the slowest RTs were observed in the very difficult filtered condition. The RT advantage for compressed speech probably reflects the fact that because these compressed words were twice as fast, complete information for word identification was available twice as fast. This indicates that the early availability of lexical information is evidently an advantage that outweighs the mild disadvantage imposed by this degree of compression. Second, straightforward developmental effects on RTs were observed only for unaltered speech. For time-compressed and low-pass-filtered words there was a trend toward developmental changes in RT only when children were grouped by vocabulary level. Altogether, it is reassuring that our findings for unaltered speech are largely congruent with a previous study by Fernald et al. (1998) in perceptually normal word stimuli using frame-by-frame coding. Overall, the developmental changes we found for RT were less robust than those observed for accuracy. We suggest three possibilities that might account for this difference. First, infants' speed in recognizing familiar spoken words was assessed by using a button-press box that inevitably introduced additional variability due to the response latency of the coder, that is, the amount of time it took the coder to press the button indicating a change in the child's direction of gaze. Using frame-by-frame coding can eliminate this variability, yielding more precise and reliable measurement of children's RT (Fernald et al., 1998). Second, RTs are calculated from distracter-initial trials only, whereas accuracy scores were calculated from both distracter- and target-initial trials. Furthermore, missing RTs in 22 children in one or more of the three perceptual conditions also reduced the population of trials that was available for assessing infants' response speed. Thus, reductions in both the number of trials per participant and the number of participants included unfortunately reduced the power in these analyses. Third, straightforward developmental effects on RTs in acoustically more challenging conditions may emerge later and thus would be evident in children beyond the ages tested here.

Contributions of Age and Vocabulary Size to Efficiency in Word Recognition

Finally, one of our major goals was to compare age and vocabulary size as converging but partially independent predictors of infant word recognition. This was addressed by grouping the same children once by age levels and then again by levels of expressive vocabulary size. Additionally, this question was extended to a larger array of stimuli and developmental levels than have been used in previous research using comparable methods. This complex experimental design yielded several important findings.

The major finding was that efficiency in word recognition increased *both* with the age of the children *and* with the size of their productive vocabulary. However, there were important differences among the three conditions of stimulus integrity. In the ANCOVAs over accuracy, vocabulary size made a unique contribution to differences in the reliability of children's responses to spoken words. In the ANCOVAs over RTs, again the only significant predictor of response speed was vocabulary size.

For unaltered words, accuracy in recognizing familiar words increased the older the children got and the more words they produced. Significant developmental changes in accuracy were evident between the 12-to-14-month and the 15-to-18-month groups, and between the 15-to-18-month and the 24-to-31-month groups. As for speed, RTs did not decrease in children

prior to the 15-month or 20-word levels. From then on children responded more quickly with increasing age and vocabulary size.

For compressed words, somewhat surprisingly, overall accuracy was only slightly lower than accuracy in unaltered words. After a reliable increase in target recognition from the 12-to-14-month group to the 15-to-18-month group, there was a gradual increase from then on, although this increase was not significant. The fact that children could recognize time-compressed words quite easily suggests that young listeners can cope with speech speeded up by 50% compared to the normal rate, with “normal” referring to infant-directed speech. In terms of response speed, a developmental trend with decreasing RTs from the 21-to-99-word group onward was evident when children were grouped by levels of vocabulary. Age groupings did not show a developmental trend.

For filtered words, the following developmental picture emerged. Expressive vocabulary size proved to be a more sensitive index of the child’s performance than was chronological age in the filtered condition. When children were grouped by age, there was no evidence of an increase in performance in recognizing low-pass-filtered words at any point from 12 to 31 months of age. However, when children were grouped by vocabulary, improved performance on filtered word trials was evident in children with 100 to 300 words in their vocabularies. Oddly, performance on filtered words then appeared to decrease in those children with the largest vocabularies. Closer inspection of the looking pattern in 300-plus-word group offered a possible explanation for this anomaly. When these children happened to start out the trial already looking at the target picture, they tended to shift away quickly, checking back at the distracter picture before returning to the correct picture later on in the trial. Because this behavior was not observed in any of the other perceptual conditions or groups, it suggests that the distracter check may have served to verify that the child was indeed fixating the correct picture. Given the poor acoustic quality of the target word, it was as if these children were saying to themselves, “Did she actually say *dog*?” As for speed, only when children were grouped by vocabulary level was there a developmental trend with a gradual decrease in RTs. When children were grouped by age, RTs stayed more or less flat across the age groups.

Why did vocabulary and age groupings produce divergent results for accuracy in this particular condition? Although age and vocabulary are highly correlated, when children were grouped by age levels, some children with higher and lower vocabularies ended up in the same pool. Correct performance of children with higher vocabularies may have been concealed by random performance of children with lower vocabularies. It seems that with increasing processing “costs,” as in cases of spectral reduction, resilience in the face of more taxing listening conditions can be better predicted by expressive vocabulary than by chronological age. However, target recognition in low-pass-filtered words was much lower than that obtained in unaltered and compressed words in the same children. Children in the 100-to-300-word group improved in accuracy on filtered word trials as compared to younger infants. Although the difference between the 100-to-300 group and the 300-plus group was not significant, there was evidence that the lexically most advanced children improved in accuracy as well. This suggests a link between speech production and online word comprehension showing that children who have more advanced lexical skills are less vulnerable to challenges in the speech signal. Links between speech production and comprehension have been documented earlier through observational studies reporting a correlation between number of words produced and number of words comprehended in the 2nd year (Bates, Bretherton & Snyder, 1988) as well as more recently through online word recognition suggesting that children with higher vocabularies are also more efficient in speech processing (Fernald et al., 2001; Fernald, Perfors, & Marchman, in press). This study

supports these results and extends the connection between production and comprehension to the domain of acoustic manipulation.

We suggest three possibilities that may at least partly explain the fact that the recognition of low-pass-filtered words was generally more advanced in higher vocabulary groups. First, this result may relate to the connection between expressive vocabulary size and strength of lexical representations. Generally, little is known to what extent lexical representations are specified in young learners. Although it was previously assumed that young learners have only vague representations of words (Walley, 1993), more recent research has suggested that children as young as 14 months may already have well-specified lexical representations for familiar words (Swingley & Aslin, 2000, 2002). Werker and colleagues (2002), however, suggested that when learning words young children may not be able to use all the phonetic detail that is perceptually available to them because of the high computational demands involved in word learning. According to this reasoning, language novices fail to establish a correct match between word form and its meaning because they simply run out of attentional resources. Older, more linguistically experienced children on the other hand may use attentional resources more efficiently and, as a result, successfully link a newly heard word with a new object.

Although this research does not allow us to directly address the question of how lexical representations are specified in young learners, the more advanced performance of children with higher vocabularies in recognizing filtered words may indicate a relationship between vocabulary size and strength of lexical representations in a more challenging listening condition. Children who produce more words may simply be better in recognizing spectrally poor words because they have more mature, more robust lexical representations and may require less spectral information for a word to recognize it. However, children with fewer words may have less solid lexical representations that are more vulnerable to “added challenge,” requiring more acoustic information to recognize a word. Thus, in this account only strongest representations “survived” and enabled the activation of the appropriate response. This reasoning is in line with a “graded representations” account proposed by Munakata (2001). According to this account, the acquisition of knowledge is viewed as graded rather than all-or-none, which means that learners may succeed in less demanding but fail in more demanding conditions when performing a task. In this context this then suggests that weaker lexical representations may be sufficient for success in word recognition in the less demanding normal and compressed conditions. However, success in the more challenging task of recognizing filtered words may require stronger representations. Second, it is also possible that word frequency could at least partly explain the result, an explanation that does not exclude the account just offered. Perhaps children with higher vocabularies have had more experience with the words tested than have children with lower vocabularies. Along these lines, more experience may have helped to identify a phonetically vaguely specified word. Finally, children with higher vocabularies may show better performance in filtered words because of developmental differences in nonlinguistic cognitive capacities that may be necessary to cope with the more challenging processing conditions of filtered words in this task (see Fernald et al., 2001). According to this logic, one possibility is that children with higher vocabularies detect and encode low-level perceptual information more efficiently and rely on more advanced attentional resources, which may allow a more successful mapping of the word form and its meaning (see Werker et al., 2002).

These various explanations all support the point that success in word recognition in more challenging listening conditions may depend on a complex interplay of linguistic as well as nonlinguistic cognitive factors, all of which may be involved in the rapid and reliable recognition of familiar words spoken under different conditions of stimulus integrity.

However, the size of the expressive vocabulary unambiguously plays an important role. Although we have data only on children's productive vocabulary size at a particular time and not on their rate or pattern of vocabulary acquisition, it is still interesting to note that the ability to recognize filtered words emerged around the time that children are reported to experience an acceleration in their lexical development, commonly referred to as the "vocabulary spurt" (Fenson et al., 1994).

Swingley and Aslin (2000) reported that effects of word-initial mispronunciations were not related to age or to spoken vocabulary size in their study. However, several studies have documented a relationship between children's vocabulary size and their speech-processing abilities in this age range (Bates & Goodman, 1997; Fernald et al., 2001; Mills, 1999; Mills et al., 1997). In a recent longitudinal study, Fernald et al. (in press) found that measures of speed and accuracy in online word recognition by 25-month-old children were consistently correlated with receptive and productive vocabulary measures at earlier ages. A relationship between vocabulary size and spoken-word recognition has also been reported recently in older children (3–8 years of age; Munson, 2001). Using two spoken-word recognition tasks—gated words (where the final stop in a consonant-vowel-consonant-combination was deleted) and noise-center words (where the initial vowel was replaced by broadband noise)—Munson demonstrated that receptive and productive vocabulary size were reliable predictors of a significant proportion of variance in children's ability to recognize spoken words under challenging conditions. Interestingly, vocabulary was a better predictor than were other measures such as preliteracy skills, phonological awareness, and articulation accuracy. The fact that vocabulary size proved to be a powerful predictor of success in word recognition suggests to us that there are major changes in the efficiency of lexical processing emerging as expressive vocabulary increases (see Reznick & Goldfield, 1992).

CONCLUSION

In summary, it is impressive that infants just learning words and building their grammar can identify words in a variety of acoustic contexts. This ability, however, is still rudimentary and the child's efficiency can be mildly disrupted by speeding up the respective target and more severely disrupted by reducing the amount of spectral information. Independent of the magnitude of the disruption, this study suggests that children's early word comprehension skills are strong enough to withstand temporal as well as spectral manipulations within certain limits. Most importantly, the results indicate that children develop higher resilience to acoustic challenges such as time compression and low-pass filtering with increasing language experience. Furthermore, developmental profiles in general seem to be better predicted by levels of vocabulary than by levels of age. This finding also illuminates the connection between receptive skills, measured as the information about the target word unfolds in real time, and productive skills, measured by parental report: There was evidence suggesting that more advanced expressive lexical development might be related to greater resilience when confronted with spectral degradation in an online word-recognition task. The fact that infants in their 2nd year of life are able to establish a connection between an acoustically modified sound form and the appropriate visual referent clearly demonstrates that efficiency in speech processing is well underway by then. We note in concluding that such an approach, emphasizing the "strength" or "shakiness" of linguistic knowledge as a factor in language-related behavior, has also proven useful in developmental studies of nonlinguistic cognition from the 1st year of life (Munakata, McClelland, Johnson, & Siegler, 1997). Hence this approach to early word understanding may be part of a unified theory of cognitive development in which knowledge and efficiency are two aspects of the same processes of development and change.

Acknowledgments

Renate Zangl is now at the Department of Psychology, Stanford University; Lindsay Klarman is now at the University of Washington, Seattle.

This work was supported by post-doctoral Grants to Renate Zangl (J-1646, J-1817) from the Austrian Science Foundation, and by a Grant from NIDCD, "Origins of Communicative Disorders" (DC 01289-0351), to Elizabeth Bates.

Our thanks to Judy Weir and Robert Buffington for technical support, to Lin Chang, Ramona Friedman, Bryce Pearson, Leah Shuchter, Alexandra Wydogen and Sheri Yoshimi for assistance in data collection and to Sriram Balasubramanian for help in editing. We are especially grateful to Kim Plunkett, Graham Schafer, Alycia Cummings and Amy Perfors for detailed advice at various stages of the project. Special thanks also to Melissa Schweisguth for her help in the early stages of the project as well as the collection and analyses of adult pilot data. Special thanks also to Virginia Marchman for helpful comments on earlier versions of the manuscript. Above all, we would like to thank the parents and their children for their participation in the present research.

REFERENCES

- Aydelott J, Bates E. Effects of acoustic distortion and semantic context on lexical access. *Language and Cognitive Processes*. 2004; 19:29–56.
- Bates, E.; Bretherton, I.; Snyder, L. From first words to grammar: Individual differences and dissociable mechanisms. Cambridge, England: Cambridge University Press; 1988.
- Bates E, Goodman J. On the inseparability of grammar and the lexicon: Evidence from acquisition, aphasia and real-time processing. *Language and Cognitive Processes*. 1997; 12:507–586.
- Blumstein SE, Katz B, Goodglass H, Shrier R, Dworetzky B. The effects of slowed speech on auditory comprehension in aphasia. *Brain and Language*. 1985; 24:246–265. [PubMed: 2579705]
- Campbell T, McNeil M. Effects of presentation rate and divided attention on auditory comprehension in children with an acquired language disorder. *Journal of Speech and Hearing Research*. 1985; 28:513–520. [PubMed: 4087886]
- Caselli MC, Bates E, Casadio P, Fenson J, Fenson L, Sanderl L, et al. Across-linguistic study of early lexical development. *Cognitive Development*. 1995; 10:159–199.
- Caselli MC, Casadio P, Bates E. A comparison of the transition from first words to grammar in English and Italian. *Journal of Child Language*. 1999; 26:69–111. [PubMed: 10217890]
- Caselli, MC.; Casadio, P.; Bates, E. Lexical development in English and Italian. In: Tomasello, M.; Bates, E., editors. *Language development: The essential readings*. Malden, MA: Blackwell; 2001. p. 76-110.
- Dick F, Bates E, Wulfeck B, Utman J, Dronkers N, Gernsbacher M. Language deficits, localization and grammar: Evidence for a distributive model of language breakdown in aphasics and normals. *Psychological Review*. 2001; 108:759–788. [PubMed: 11699116]
- Dorman, MF.; Loizou, PC.; Kirk, KL.; Svirsky, M. Channels, children and the Multisyllabic Lexical Neighbourhood Test (MLNT); Paper presented at the National Institutes of Health Neural Prosthesis Workshop; Bethesda, MD. 1998 October.
- Dupoux E, Green K. Perceptual adjustment to highly compressed speech: Effects of talker and rate changes. *Journal of Experimental Psychology: Human Perception and Performance*. 1997; 23:914–927. [PubMed: 9180050]
- Eisenberg LS, Shannon RV, Schaefer Martinez A, Wygonski J. Speech recognition with reduced spectral cues as a function of age. *Journal of the Acoustical Society of America*. 2000; 107:270–2710.
- Fenson L, Dale PA, Reznick JS, Bates E, Thal D, Pethick SJ. Variability in early communicative development. *Monographs of the Society for Research in Child Development*. 1994; 59(5 Serial No. 242)
- Fenson, L.; Dale, PA.; Reznick, JS.; Thal, D.; Bates, E.; Hartung, J., et al. *MacArthur Communicative Developmental Inventories: User's guide and technical manual*. San Diego, CA: Singular Publishing Group; 1993.

- Fernald A, Kuhl PK. Acoustic determinants of infant preference for motherese speech. *Infant Behavior and Development*. 1987; 10:279–293.
- Fernald A, Perfors A, Marchman V. Picking up speed in understanding: How increased efficiency in on-line speech processing relates to lexical and grammatical development in the second year. *Developmental Psychology*. (in press).
- Fernald A, Pinto JP, Swingley D, Weinberg A, McRoberts GW. Rapid gains in speed of verbal processing by infants in the 2nd year. *Psychological Science*. 1998; 9:228–231.
- Fernald A, Simon T. Expanded intonation contours in mothers' speech to newborns. *Developmental Psychology*. 1984; 20:104–113.
- Fernald A, Swingley D, Pinto JP. When half a word is enough: Infants can recognize spoken words using partial phonetic information. *Child Development*. 2001; 72:1003–1015. [PubMed: 11480931]
- Fernald A, Taeschner T, Dunn J, Papousek M, de Boysson-Bardies B, Fukui I. A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants. *Journal of Child Language*. 1989; 16:477–501. [PubMed: 2808569]
- Foulke, E. The perception of time compressed speech. In: Horton, DL.; Jenkins, JJ., editors. *The perception of language*. Columbus, OH: Merrill; 1971. p. 79-101.
- Golinkoff RM, Hirsh-Pasek K, Cauley KM, Gordon L. The eyes have it. Lexical and syntactic comprehension in a new paradigm. *Journal of Child Language*. 1987; 14:23–45. [PubMed: 3558524]
- Gordon-Salant S, Fitzgibbons PJ. Profile of auditory temporal processing in older listeners. *Journal of Speech, Language and Hearing Research*. 1999; 42:300–311.
- Grosjean F. The recognition of words after their acoustic offset: Evidence and implications. *Perception & Psychophysics*. 1985; 38:299–310. [PubMed: 3831907]
- Haith MM, Wentworth N, Canfield RL. The formation of expectations in early infancy. *Advances in Infancy Research*. 1993; 8:251–297.
- Hollich GJ, Hirsh-Pasek K, Golinkoff RM, Brand RJ, Brown E, Chung HL, et al. Breaking the language barrier: An emergentist coalition model for the origins of word learning. *Monographs of the Society for Research in Child Development*. 2002; 65(3) v–123.
- King PE, Behnke RR. The effect of time-compressed speech on comprehension, interpretive, and short-term listening. *Human Communication Research*. 1989; 15:428–443.
- Lane H, Grosjean F. Perception of reading rate by speakers and listeners. *Journal of Experimental Psychology*. 1973; 97:141–147. [PubMed: 4705233]
- Leonard CL, Baum SR, Pell MD. Context use by right-hemisphere-damaged individuals under a compressed speech condition. *Brain and Cognition*. 2000; 43:315–319. [PubMed: 10857716]
- Manning W, Johnston K, Beasley D. The performance of children with auditory perceptual disorders on a time-compressed speech discrimination measure. *Journal of Speech and Hearing Disorders*. 1977; 42:77–84. [PubMed: 839757]
- Marchman V, Bates E. Continuity in lexical and morphological development: A test of the critical mass hypothesis. *Journal of Child Language*. 1994; 21:339–366. [PubMed: 7929685]
- Marslen-Wilson, WD. Function and process in spoken word recognition: A tutorial review. In: Bouma, H.; Bouwhuis, D., editors. *Attention and performance X: Control of language processes*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.; 1984. p. 125-148.
- Marslen-Wilson WD. Functional parallelism in spoken word-recognition. *Cognition*. 1987; 25:71–102. [PubMed: 3581730]
- Marslen-Wilson, WD. Issues of process and representation in lexical access. In: Altmann, GT.; Shillcock, R., editors. *Cognitive models of speech processing: The second Sperlonga meeting*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.; 1993. p. 187-211.
- McNutt J, Chi-Yen Li J. Repetition of time-altered sentences by normal and learning-disabled children. *Journal of Learning Disabilities*. 1980; 13:25–29. [PubMed: 7373142]
- Mills, DL. Cerebral specialization before and after the vocabulary spurt: An event-related potential study of novel word learning. *Early vocabulary growth: Mechanisms of change; Symposium conducted at the Biennial Meeting for Research in Child Development; Albuquerque, NM. 1999 April.*

- Mills DL, Coffey-Corina SA, Neville HJ. Language acquisition and cerebral specialization in 20-month-old infants. *Journal of Cognitive Neuroscience*. 1993; 5:317–334.
- Mills DL, Coffey-Corina SA, Neville HJ. Language comprehension and cerebral specialization from 13 to 20 months. *Developmental Neuropsychology*. 1997; 13:397–445.
- Mills DL, Pratt C, Zangl R, Stager C, Neville H, Werker J. Language experience and the organization of brain activity to phonetically similar words: ERP-evidence from 14- and 20-month-olds. *Journal of Cognitive Neuroscience. Special Issue on Developmental Neuroscience*. 2004; 16:1452–1464.
- Munakata Y. Graded representations in behavioral dissociations. *Trends in Cognitive Science*. 2001; 5:309–315.
- Munakata Y, McClelland JL, Johnson MH, Siegler R. Rethinking infant knowledge: Toward an adaptive process account of successes and failures in object permanence tasks. *Psychological Review*. 1997; 104:686–713. [PubMed: 9337629]
- Munson B. Relationships between vocabulary growth and spoken word recognition in children aged 3–7. *Contemporary Issues in Communication*. 2001; 28:20–29.
- Naigles L. First language acquisition: Method, description, & explanation. *Language and Speech*. 1990; 33:175–180.
- Nooteboom, SG.; Doodeman, GJN. Speech quality and the gating paradigm. In: van den Broek, MPR.; Cohen, A., editors. *Proceedings of the Tenth International Congress of Phonetic Sciences*; Dordrecht, Netherlands: Foris; 1984. p. 48–485.
- Remez RE, Rubin PE, Pisoni DB, Carrell TD. Speech perception without traditional speech cues. *Science*. 1981 May 22;212:947–950. [PubMed: 7233191]
- Reznick JS, Goldfield BA. Rapid change in lexical development in comprehension and production. *Developmental Psychology*. 1992; 28:406–413.
- Schafer G, Plunkett K. Rapid word learning by fifteen-month-olds under tightly controlled conditions. *Child Development*. 1998; 69:309–320. [PubMed: 9586207]
- Schmitt JF, Carroll MR. Older listeners' ability to comprehend speaker-generated rate alteration of passages. *Journal of Speech and Hearing Research*. 1985; 28:309–312. [PubMed: 4010262]
- Shannon RV, Zeng F-G, Wygonski J. Speech recognition with altered spectral distribution of envelope cues. *Journal of the Acoustical Society of America*. 1998; 104:2467–2476. [PubMed: 10491708]
- Shannon RV, Zeng F-G, Wygonski J, Kamath V, Ekelid M. Speech recognition with primarily temporal cues. *Science*. 1995 October;270:303–304. [PubMed: 7569981]
- Speer SR, Wayland SC, Kjelgaard MM, Wingfield A. Effect of speaking rate and paragraph structure on the production of sentence prosody. *Journal of the Acoustical Society of America*. 1994; 95:2979.
- Stine EAL, Wingfield A, Myers SD. Age differences in processing information from television news: The effects of bisensory augmentation. *Journal of Gerontology: Psychological Sciences*. 1990; 45:P1–P8.
- Swingle D, Aslin D. Spoken word recognition and lexical representation in very young children. *Cognition*. 2000; 76:147–166. [PubMed: 10856741]
- Swingle D, Aslin D. Lexical neighbourhoods and the word-form representations of 14-month-olds. *Psychological Science*. 2002; 13:480–484. [PubMed: 12219818]
- Swingle D, Fernald A. Recognition of words referring to present and absent objects in 24-month-olds. *Journal of Memory and Language*. 2002; 46:39–56.
- Swingle, D.; Pinto, JP.; Fernald, A. Assessing the speed and accuracy of word recognition in infants. In: Rovee-Collier, C.; Lipsitt, LP.; Hayne, H., editors. *Advances in infancy research*. Vol. Vol. 12. Stamford, CT: Ablex; 1998. p. 257–277.
- Swingle D, Pinto JP, Fernald A. Continuous processing in word recognition at 24 months. *Cognition*. 1999; 71:73–108. [PubMed: 10444905]
- Tanenhaus MK, Spivey-Knowlton MJ, Eberhard KM, Sedivy JE. Integration of visual and linguistic information in spoken language comprehension. *Science*. 1995 June;268:1632–1634. [PubMed: 7777863]

- Turner CW, Souza PE, Forget LN. Use of temporal envelope cues in speech recognition by normal and hearing-impaired listeners. *Journal of the Acoustical Society of America*. 1995; 97:2568–2576. [PubMed: 7714274]
- Walley AC. The role of vocabulary development in children's spoken word recognition and segmentation ability. *Developmental Review*. 1993; 13:286–350.
- Weismer SE, Hesketh LJ. Lexical learning by children with specific language impairment: Effects of linguistic input presented at varying speaking rates. *Journal of Speech and Hearing Research*. 1996; 39:177–190. [PubMed: 8820709]
- Werker JF, Fennell CT, Corcoran KM, Stager CL. Infants' ability to learn phonetically similar words: Effects of age and vocabulary size. *Infancy*. 2002; 3():1–30.
- Werker, JF.; Stager, CL. Developmental changes in infant speech perception and early word learning: Is there a link?. In: Broe, M.; Pierrehumbert, J., editors. *Papers in Laboratory Phonology*. Vol. 5. Cambridge, England: Cambridge University Press; 2000. p. 18-193.
- Wingfield A, Tun PA, Koh CK, Rosen MJ. Regaining lost time: Adult aging and the effect of time restoration on recall of time-compressed speech. *Psychology and Aging*. 1999; 14:380–389. [PubMed: 10509694]
- Wingfield A. Cognitive factors in auditory performance: Context, speed of processing, and constraints of memory. *Journal of the American Academy of Audiology*. 1996; 7:175–182. [PubMed: 8780990]

FIGURE 1.

Accuracy across two 1-sec time windows. Proportional accuracy is plotted separately for normal, time-compressed, and low-pass-filtered words for each accuracy window. The vertical axis indicates the children's proportion of target fixations; the horizontal axis indicates the two time windows (first second = first 1-sec time window; second second = second 1-sec time window).

FIGURE 2.

Accuracy grouped over Age Levels—Second Time Window. Proportional accuracy is plotted separately for normal, time-compressed, and low-pass-filtered words. The vertical axis indicates the children's proportion of target fixations; the horizontal axis indicates the four age levels.

FIGURE 3.

Accuracy grouped over expressive Vocabulary Levels—Second Time Window. Proportional accuracy is plotted separately for normal, time-compressed, and low-pass-filtered words. The vertical axis indicates the children's proportion of target fixations; the horizontal axis indicates the four vocabulary levels.

TABLE 1

Number of Infants for Each Age, as it is Currently

Age (in Months)	Number of Infants	Age (in Months)	Number of Infants
12	<i>N</i> = 13	21	<i>N</i> = 6
13	<i>N</i> = 6	22	<i>N</i> = 6
14	<i>N</i> = 3	23	<i>N</i> = 4
15	<i>N</i> = 7	24	<i>N</i> = 11
16	<i>N</i> = 8	25	<i>N</i> = 2
17	<i>N</i> = 3	26	<i>N</i> = 1
18	<i>N</i> = 7	27	<i>N</i> = 1
19	<i>N</i> = 11	30	<i>N</i> = 2
20	<i>N</i> = 3	31	<i>N</i> = 1

TABLE 2

List of 24 Target Words Used in This Study

Target Word	Target Word	Target Word
ball	cup	hat
bed	dog	horse
bird	doll	keys
book	door	mouth
car	ear	nose
cat	eyes	phone
chair	foot	pig
cow	hand	shoe

TABLE 3

Cross-tabulation of Expressive Vocabulary Size by Age Groupings

Age	12–14 Months	15–18 Months	19–23 Months	24–31 Months
0–100 words	17	8	2	0
101–200 words	5	11	11	4
201–300 words	0	6	13	2
more than 300 words	0	0	4	12