On-line word processing in infants and toddlers:
The role of age, vocabulary and perceptual degradation

Renate Zangl*
Lindsay Klarman**
Donna Thal***
Elizabeth Bates*

* Center for Research in Language, University of California, San Diego,
** University of Washington, Seattle, *** San Diego State University

Technical Report CND-0101
2001
Project in Cognitive and Neural Development
Center for Research in Language
University of California, San Diego
La Jolla, CA 92093-0526

Abstract

The effects of acoustic degradation on lexical processing were investigated in children ranging from 12 to 31 months of age. Using a visual fixation technique, familiar target words (in infant-directed speech) were presented either acoustically unaltered, time compressed (50%) or low-pass filtered (1.5kHz). The children's ability to correctly identify the respective target was assessed by examining (1) accuracy, (2) peak latency (= response time) and (3) peak duration (as a measure of visual engagement). All three factors were sensitive to the acoustic manipulations; the severity of the effect was dependent on the nature of the acoustic distortion, the child's age and—more importantly—her vocabulary level. Adults tested using identical and more severe perceptual degradations mirrored the results obtained in children in their overall pattern but differed in the magnitude of the effects. Low-pass filtering had minimal effects on word recognition in adults, but the same manipulation had devastating effects on infants, with accurate looking observed only in the more sophisticated groups (vocabularies >100 words). In contrast, infants had relatively little difficulty recognizing temporally compressed stimuli, although they did show enhanced stimulus engagement for compressed compared with unaltered words (i.e. "staring longer"). Reaction time data indicate a sharp, non-linear increase in the efficiency of word recognition that coincides with the well-known acceleration in expressive vocabulary known as the "vocabulary burst".

Introduction

Comprehending spoken language is a complex cognitive task that works at high speed and requires precise timing in the integration of multiple sources of information from the incoming speech stream. Adult listeners process language incrementally, rapid-ly merging linguistic and contextual information from the speech signal (Marslen-Wilson, 1984, 1987, 1993). A skilled listener can handle around 10-15 phonemes per second, cope easily with 140 to 180 words arriving per minute, overcome coarticulation conflicts, adjust to variabilities in talkers, contin-uously update the interpretation of the running speech signal, and extract meaning before the actual off-set of the acoustic event (Lane & Grosjean, 1973, Speer, Wayland, Kjelgaard, & Wingfield, 1994, Stine, Wingfield & Myers, 1990). Correct identification of words in a sentence context is performed around 200 ms post-stimulus onset; word recognition in isolation is performed, on average, within an additional 130 ms (Grosjean 1980, Marslen-Wilson, 1984, Tyler 1984). The astonishing efficiency and optimal use of speech processing strategies in adults may sometimes conceal the enormous precision and "teamwork" that is required to decode spoken language correctly. In fact, more taxing processing climates affecting signal clarity—such as background noise, simultaneous presentation of speech from different spatial sources, accelerated speaking rate, reduced spectral information—can put the comprehension of spoken language at risk (Grosjean 1985; Nooteboom & Doodeman, 1984; Dick, Bates, Wulfeck, Utman, Dronkers, & Gernsbacher; Utman & Bates, 2000). If speech is acoustically distorted, processing demands increase and the process of language comprehension can be disrupted or delayed.

Although perceptual degradation can lead to decrease in performance (depending on the severity of distortion), adult word recognition is usually robust and efficient. A series of recent studies in infant word recognition provide evidence that infant listeners are also efficient at processing words by the end of the second year, for acoustically unaltered speech. However, the impact of perceptual degradation on the activation of linguistic representations in the early stages of language learning is largely unexplored (for an important exception, see Swingley & Aslin, 2000, discussed below). In contrast to adults, infants are linguistically less experienced and have a much shakier system to start with. The well-documented use of infant-directed speech in early infant-interaction may serve to reduce stress on the infant's fragile system. But under acoustically more demanding conditions, infants may experience decrements in processing efficiency that are much more severe than those observed for adults in similar conditions.

The goal of the present study was to investigate the real time dynamics of early word processing between the ages of 12 and 31 months. Using the preferential looking technique, familiar auditory word targets were presented under three different conditions: (1) acoustically unaltered, (2) time-compressed, and (3) low-pass filtered. In order to fine-tune the level of degradation to the infant's system, we first tested adults using various degrees of degradation. Including adults not only provided the opportunity to assess single word identification in a mature, more efficient system but also to compare their overall performance patterns with the ones obtained in infants. In other words, it allowed us to investigate whether infants actually behave like adults, exhibiting differences 'only' in the magnitude of the effects (= quantitative differences), or whether they behave differently, showing distinct patterns not encountered in adults (= qualitative differences).

Temporal compression was manipulated to investigate developmental changes in speed of word processing; low-pass filtering was manipulated to assess the amount and quality of acoustic information that infants must have to recognize a word and achieve a word-picture match. To our knowledge, there is no prior work on the effects of these 'stressors' on infant processing; hence the present study is necessarily exploratory in nature. To create compressed and filtered word stimuli, we selected parameters that were shown (in pretests) to have a measurable but relatively small effect on processing in adults. Furthermore, we approached the question of developmental effects on word recognition and 'resistance to stress' from two points of view: comparisons over age, and comparisons over levels of expressive vocabulary. To the extent that efficiency of word comprehension reflects maturational factors, we might expect to find effects of age that are independent of vocabulary level. On the other hand, as we will see in more detail below, 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; Marchman & Bates 1994, Mills et al. 1993, Mills et al. 1997; Mills et al. (under review), Fernald et al. 1998, Fernald et al. 2001, Swingley, Pinto & Fernald 1999, Munson in press). We selected the age range between 12 to 31 months because the most dramatic developmental changes in language take place during this period. The first systematic evidence for word comprehension usually appears in naturalistic contexts between 8-10 months of age, but receptive vocabulary accelerates sharply from 12-16 months (Fenson et al. 1994). The first systematic evidence of word production is usually reported around 12 months of age (Bates, Bretherton & Snyder, 1988). After a slow start, rate of expressive vocabulary growth accelerates markedly between 18-20 months for most children, a phenomenon often referred to as the "vocabulary burst" (Goldfield & Reznick, 1990, 1996, Mervis & Bertrand 1995, Bloom 1993, Roberts 1998). This surge in expressive vocabulary co-occurs with the emergence of word combinations, and further developments in grammar are tightly correlated with vocabulary size up to 30/31 months of age – the upper limit in the present study (Bates & Goodman, 1997).

Considering the huge increase in expressive vocabulary and grammar that takes place across the 12-31 month age range, the question arises whether and to what extent there are similar developmental changes in the infant's receptive processing skills across the same period. A number of recent on-line studies (using preferential looking or event-related potentials) have greatly enhanced our knowledge of auditory language processing across this period. The results suggest a rapid increase in the efficiency of lexical processing, involving increases in accuracy, decreases in reaction time, accompanied by a high sensitivity towards acoustic phonetic deviations and contextual cues from around 18-20 months onwards (Fernald, Pinto, Swingley, Weinberg & McRoberts 1998, Schafer & Plunkett 1998, Swingley, Pinto & Fernald 1999, Swingley & Aslin 2000, Fernald, Swingley & Pinto 2001). At the same time, changes in brain organization have been reported between 13 and 20 months of age which seemed to be linked to the infant's vocabulary size even after age is controlled (Mills et al. 1993, Mills et al. 1997, Mills 1999).

Within the context of early word recognition, infants have been shown to need only partial word information to activate correct lexical representations (*/bei/ instead of baby*) and to be able to rely on contextual information provided in sentences/phrases in early picture-word mapping (Fernald 2001). Furthermore, Swingley and Aslin (2000) provided evidence that infants do in fact have phonetically well-specified representations of familiar words instead of phonetically vague, less detailed representations as previously suggested (Walley 1993, Jusczyk 1986, Charles-Luce & Luce 1990). In using good exemplars of familiar words and mispronounced counterparts deviating in the initial segment (*baby – vaby)* they demonstrated that infants were sensitive to the mispronunciations. Although the children were still able to recognize the mispronounced target, their performance decreased both in accuracy and reaction time. Further evidence for incremental processing in the later stages of the second year has also been shown through the use of event-related potentials (ERPs). Mills, Pratt, Stager, Zangl, Neville & Werker (in preparation) compared components of the ERPs to familiar words (e.g. *bear*), phonetically similar nonsense words (constructed by replacing the initial phoneme in a known word, e.g. *gare*), and phonetically dissimilar nonsense words (e.g. *lif*). In 20-month-olds, the ERPs were different for real words vs. phonetically similar nonsense words, suggesting that words are fully specified at the phonetic level. In contrast, the ERPs in 14-month-olds were similar for real words and phonetically similar nonsense words (e.g. no difference between *bear* and *gare*, but both differed from *lif*), suggesting that phonetic representations for familiar words may be underspecified at this age. A recent preferential looking study by Swingley, Pinto and Fernald (1999) shows that 24-month-olds are able to exploit the internal phonetic structure of words, using the first phoneme to anticipate a word-picture match. For example, given a choice between two pictures whose names begin with a different phoneme (e.g. *dog – tree*), infants begin to look toward the appropriate target well before word offset. By contrast, given two target pictures whose names begin with the same phone (e.g. *dog—doll*), infants did not move their eyes toward the correct picture until some point close to the end of the word.

These studies suggest that infants process the incoming speech stream continuously, can identify words with incomplete signal specification, can merge linguistic and non-linguistic information, and can exploit temporal characteristics in word processing in ways that are qualitatively similar to word processing in adults. Most important for the orientation of the present study, the results suggest that the infant's performance and efficiency are sensitive to the specific acoustic-phonetic characteristics of the speech signal, reflected by differential processing in situations of increased processing demands such as in phonemic overlap at word-onsets or mispronunciations. These results set the stage for the present study, in which developmental changes in word processing are compared under normal, temporally compressed and spectrally filtered conditions.

What are the effects of time compression and low-pass filtering on the acoustic signal itself? Temporal compression algorithms increase the rate of the original speech signal by periodically deleting small segments of the signal at regular intervals. Because it preserves crucial spectral information and retains normal rhythmic and prosodic information, time-compressed speech sounds quite natural to the 'naked ear'. Although the resulting signal is quite intelligible (up to relatively high amounts of compression), the listener is given substantially less time to recognize words and integrate them into the current context. For this reason, time compression is believed to affect "top-down", central aspects of language processing (Gordon-Salant & Fitzgibbons 1993, Connolly et al. 1990). For example, Utman & Bates (2000) have shown that compressed speech results in a reduction in listener's ability to suppress word candidates that are unrelated to the context.

Low-pass filtering algorithms remove high-frequency spectral information that is important for many segmental speech contrasts. However, it retains the prosodic and temporal aspects of the speech signal, and provides (up to a point) enough spectral information to permit word identification (especially in context). Informally, low-pass filtered speech resembles speech through walls of varying thickness, depending on the amount of filtering that is applied. Because the listener is working with less information (although the amount of time required to process that information is unchanged), low-pass filtering is believed to have its greatest effects on the "bottom-up" flow of information from the speech periphery (Stuart & Phillips, 1996). For example, Utman & Bates (2000) have shown that low-pass filtering of the sentence context results in a significant reduction in priming, that is, a reduction in the facilitative effects of context on recognition of a contextually appropriate target word.

Both of these stressors push the system to its limits, but their effects differ in ways that have important implications for early language development. Almost all studies employing speech compression have been carried out with adults. This includes studies of older adults and/or neurologically impaired populations, as well as studies in which hypothesized processing deficits linked to aging or brain injury are 'simulated' by testing young adults under adverse processing conditions (Blumstein et al., 1985; Dick et al., in press; Foulke 1971, Wingfield, 1996, Leonard et al. 2000). Studies that employ time compression yield several conclusions. First, lower compression rates yield higher performance in auditory comprehension and recall abilities than higher compression rates, requiring less perceptual adaptation (Dupoux & Green 1997, Mehler et al. 1993). Older adults are significantly more vulnerable than younger adults in accelerated speech conditions (even if pure tone acuity is held constant between the groups) (Schmitt & Carroll, 1985, Gordon-Salant & Fitzgibbons 1993, 1999, Tun, Wingfield, Stine & Mecsas 1992, Tun 1998, Wingfield, Tun, Koh & Rosen 1999), and older adults are also reported to have a listening preference for slower speech rates compared to younger adults (Obler, Fein, Nicholas & Albert, 1991, Wingfield & Ducharme, 1999, Vaughan & Letowski 1997). Slower speech rate advantages were reported at word- (Sticht & Gray 1969), sentence repetition- (Stine et al. 1986) as well as discourse levels (King & Behnke, 1989). Poorer comprehension was also demonstrated in complex time-compressed syntactic constructions (Dick et al., 2001). Age effects were found at different compression rates ranging from 30% to 60% (Gordon-Salant & Fitzgibbons 1993). Recent functional imaging data indicate that the brain activation is sensitive to differences in speaking rate. Higher compression levels in sentences resulted in additional recruitment of brain regions compared to sentences at normal speaking rate (Poldrack et al. 1998). Furthermore, accelerations in speaking rate seem to be modality-independent. A recent study by Fischer, Delhorne & Reed (1999) using rate variations in American Sign Language has shown that rate effects in the visual modality roughly match those obtained in the auditory domain. A breakdown in processing emerged at increases of 2.5 to 3 times of the original rate.

A small number of studies have looked at the effects of time-compressed speech in children, with results roughly corresponding to those documented for adults. Slower rates have been reported to lead to higher production and comprehension scores in a novel word learning context (Weismer & Hesketh 1996), and comprehension and production skills are both significantly affected by speaking rate variations. Studies comparing adult-directed speech (ADS) and child-directed speech (IDS) have shown that these two registers tend to vary in speech rate (i.e. faster speech is directed to adults), suggesting that adults unconsciously decrease their speech rate to young children. Fernald and McRoberts (1991) have shown that word recognition is better for IDS than ADS in 15-month-old children, but the two registers result in equally good performance in 18-month olds. Hence there may be a developmental decrease from 15–18 months in the amount of time (and amount of information) required to recognize words. A recent study by Cooper et al. (2000) with 2- and 4-month old infants used samples of IDS which varied in rate, showing that both age groups preferred slower variants over faster ones. Hence, independent of the many other characteristics that differentiate between infant-directed and adult-directed speech, rate seems to matter.

Given this demonstrated infant preference for slower speaking rates, increased duration may have a beneficial effect in the present study, with acoustically unaltered targets potentially presenting perceptually ideal candidates for word identification. If longer duration increases auditory word comprehension, then performance scores in normal, acoustically unaltered words should be higher than those in the time compressed variants. If on the other hand, infants are equally able to identify acoustically unaltered IDS targets and their 50% compressed counterparts, no performance difference is expected with loss of processing time (Foulke & Sticht 1969). This would suggest that infants are capable of handling variations in speaking rate from very early on—at least up to a 50% compression threshold level. A third possibility is that performance in time-compressed targets varies in relation to the measures. In other words, decreased stimulus length might have different effects on accuracy, response latencies and/or visual engagement.

Studies employing low-pass filtering and related forms of acoustic degradation have also been conducted primarily with adults. In general, low-pass filtering has been shown to result in performance decrements that include increased response time and reduced accuracy (Stuart & Phillips 1996, Dick et al., 2001). A recent study by Eisenberg et al. (2000) provides evidence that younger children (5-7 years) have severe deficits in recognizing spectrally degraded speech, compared to older children (10-12 years) and adults exposed to the exact same materials and conditions (including words, sentences, syllables and digits). Adults and older children did not exhibit significant differences in performance scores; younger children, however, lagged significantly behind, requiring more spectral resolution to reach comparable performance levels in comprehension, identification and recall tasks. These results are in accordance with a previous study indicating that children aged 3 to 4 years need more spectral resolution in multisyllabic word comprehension than adults do (Dorman et al. 1998b). In infants, low-pass filtering has been extensively used in discrimination and segmentation studies investigating the role of prosodic, rhythmic versus segmental information within the first year of life (Mehler & Christophe 1995, Jusczyk et al. 1993). Its role in word recognition is currently unknown. The only study using low-pass filtering in a word-learning context had a different orientation. It examined the role of silence versus tone, word, and content-filtered word conditions on early object categorization in 9 month olds (Balaban & Waxman 1997). The results provided evidence that infants in the content-filtered word and word condition equally increased attention to novel words. This suggests, that infants in fact do pay attention to spectrally degraded stimuli. However, Cooper & Aslin (1994) demonstrated that very young infants prefer unfiltered over filtered speech, showing increased looking times during unfiltered speech. Hence this preference is in place long before the stage at which they are able to understand words.

To summarize, both types of distortions – time compression and low-pass filtering—seem to increase the demands on the listener's processing resources. Whether or not these manipulations result in perform-ance decrements seems to depend on age (Blackwell & Bates 1995, Devescovi, Pizzamiglio, Bates, Her-nandez & Marangolo, 1994, Miyake, Carpenter & Just 1994). Of particular interest for the present study is the finding that time compression and low-pass filtering may exert differential effects on the listener (Utman & Bates, 2000; Stark & Montgomery, 1995). Based on these results, we predict that perceptual degradation will affect word identification, resulting in a decrease in performance. Within this general framework, the present study had three goals.

- To compare the effects of two different forms of perceptual degradation, time compression and low-pass filtering, in experienced adult listeners (Experiment 1) and in infants in the first stages of word learning (Experiment 2).
- To determine whether developmental changes in performance within the first stages of word learning are best predicted by age, vocabulary level, or a combination of the two.

- To explore the real-time dynamics of word recognition in 12-31 month old infants, using several different measures of accuracy as well as reaction time and stimulus engagement. Within this framework, we will also compare results for different kinds of looking events or "trial conditions", including trials in which the child was looking in the wrong place at word onset (distractor or "D-trials") vs. trials in which the child was looking at the target picture at word onset (target or "T-trials").

EXPERIMENT 1

Experiment 1 focussed on adults. Adult data were collected for three reasons: to examine the effects of low-pass filtering and time compression on a more mature and skilled system; to calibrate the level of perceptual degradation (moderate vs. severe) to the infant's more fragile system; and to also be able to better evaluate the effects obtained in the first stages of word learning.

Method

Participants

Fifty-five college students (31 females, 24 males) at the University of California, San Diego participated in this experiment in exchange for academic credit. One of the participants had to be excluded due to a high number of 'no-response trials', leaving a total of fifty-four. None of the participants reported any hearing disorder and they were all between 18 and 30 years old. Participants were not familiar with the stimuli and were informed about the purpose of the experiment only after having been tested. Participants were tested in two groups depending on the level of perceptual degradation: Group A (n=27, moderate perceptual degradation) and Group B (n=27, severe perceptual degradation).

Stimuli

The entire experiment was based on 48 trials, each consisting of an auditory sentence and a pair of pictures, one of which matched the final word (= target) in the sentence.

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 a uniform off-white background on two separate 30cm color video monitors. Each auditory target had four visual exemplars, which were all prototypical instantiations of the respective word and balanced for visual salience. The images chosen were downloaded from CD Roms, the Internet, or derived from scanned digital photographs and edited on Adobe Photoshop. Each picture served twice as a target and twice as a distractor (= total of 96 pictures). Target and distractor pictures appeared simultaneously on the screens and were presented 650 ms before the onset of the sentence and, more importantly, a total of 2050 ms before the onset of the target word. The pictures stayed on through the entire auditory event and beyond; picture off-set was at 5250 ms.

Auditory stimuli

Because the study was primarily aimed at infants, the 24 target words were chosen from the earliest words comprehended by typically developing children, based on the norms of the MacArthur Communicative Developmental Inventory (Fenson et al. 1993). Table 1 provides a list of all the target words used in the experiment. The auditory stimuli were digitally recorded in a sound proof room by a female native speaker of American English, at a sampling rate of 44,000 Hz using a Sony DAT-recorder. The acoustic envelope of each word was typical of infant-directed speech, showing both extended duration and pitch patterns (Fernald et al. 1985, 1989, Cooper et al. 1990, 1993). The speech stimuli were then digitized at 22,050 Hz using Sound Designer for Macintosh and converted into 16-bit wav.files for use on a Windows/DOS system. Attention was directed to the target picture by using a standard, invariant carrier frame followed by the respective target ("Look, look at the + *target")*. The initial *Look* was identical across all sentences; 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. While the carrier frame was always presented in normal, acoustically non-modified speech, the target words had three acoustic shapes: (1) unaltered (= normal), (2) low-pass filtered, and (3) time compressed. The acoustic distortions were separately 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. *Equalizer* changed the spectral resolution of the speech signal by eliminating the frequency information at either 1.5 kHz or 1.0 kHz. *Tempo* decreased the original stimulus length by either 50% or 25% but preserved segmental and pitch information. Wave and spectrum of the target *door* in each of the three auditory conditions are presented in Figure 1.

Adults were exposed to *one* of the two degradation levels: Level A with moderate perceptual degradation (= 50% time compression and 1;5 kHz low-pass filtering); Level B with more severe perceptual degradation (= 25% time compression and 1;0 kHz low-pass filtering). The introduction of a more severe level of perceptual degradation was considered not only useful but necessary for two reasons. First, it served to externally increase the processing demand by more intense “stressors” and second, it may provide a better comparative estimate of the adult’s and children’s performance (an internally shakier system is compensated by intensified stressors).

The mean length of the unaltered/filtered stimuli was 1051.85 ms; the mean length of the compressed stimuli was 520.58 ms. Each target word was presented twice per experimental session – once unaltered and once under ONE type of acoustic distortion. The type of perceptual degradation was counterbalanced across subjects; each subject was exposed to both types of acoustic degradation. The total of 48 trials was split up into 8 blocks of 6 trials with each block containing three unaltered and three acoustically modified targets with the type of distortion being constant within but variable across blocks. The pairings of targets and distractors was based on phonological and semantic dissimilarity, meaning 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 distractors were also matched in relation to age of comprehension and age of production based on CDI norms (eg. an *easy* target had an *easy* distractor and vice versa such as *dog* and *car* vs. *pig* and *hat*).

Apparatus and Procedure

Each subject was tested individually in a sound-attenuated booth at the Center for Research in Language (University of California, San Diego). Before testing, subjects were verbally instructed to find the correct auditory-visual match by pressing either the right (=match appearing on the right side of the screen) or the left (= match appearing on the left side of the screen) button as fast and accurately as possible. The experiment was presented on a Macintosh Power PC using the Psyscope software which allowed us to control the timing of the events and provided a record of the subject’s accuracy and reaction times. If responses were not made within a 3 second time limit a “no-response” was recorded and the trial presentation advanced.

Results

The two dependent variables of interest in this study of adults were accuracy (= proportion of correct responses) and reaction time (= response time measured from the end of the sentence/off-set of the target word). Both variables were obtained for each subject individually for each of the three auditory contexts the target was presented in (= unaltered, time compressed and low-pass filtered). Mean proportional accuracy scores were calculated by dividing the number of correct responses by the total number of correct and incorrect responses for each perceptual condition. Mean RTs were obtained by averaging correct response times—again separately for each perceptual condition. Individual accuracy scores were entered into a 2 (Degree of manipulation) × 3 (Perceptual Condition) ANOVA. Similarly, individual RT scores were entered into a 2 (Degree of manipulation) × 3 (Perceptual Condition) ANOVA.

Accuracy

A 2 × 3 mixed ANOVA yielded a main effect of Perceptual Condition (only). There was no main effect of Degree of Manipulation or any interaction. As for Perceptual Condition, subjects were significantly less accurate in low-pass filtered than in time-compressed ($p = .03$) or unaltered ($p = .002$) conditions. Time-compressed and unaltered words did not differ significantly from one another ($p = .3$). Although lower than the compressed (98%) and unaltered (98.9%) condition, low-pass filtering was still very robust, reaching a mean of 96.2%. Thus, it is safe to assume that adults did not have any difficulties in correct target identification even with more severe degradation levels.

Reaction Time

A 2 × 3 mixed ANOVA revealed main effects of Degree of Manipulation (moderate vs severe; $F(1, 26) = 10.93$, $p = .002$) and of Perceptual Condition ($F(2, 52) = 261.06$, $p = .0001$).

With regard to the main effect of Degree of Manipulation, severely distorted targets had positive RTs (mean = 4.44 ms) while moderately distorted targets had negative RTs (= the subjects responded before the end of the target word; mean= –154.82 ms). The severity of manipulations obviously affected the subject’s latency with more severe distortions yielding longer latencies.

With regard to the main effect of Perceptual Condition, results showed significant differences between each of the three perceptual conditions (at the $p<.0001$ level). Time-compressed targets required more time for correct identification (mean = 176.76 ms) than did low-pass filtered targets (mean = –52.72 ms). Fastest responses were obtained in unaltered targets (mean = –349.61ms). Note that minus-responses indicate that the subjects could correctly identify the target before all of its phonetic information was available. Additionally, there were significant differences for both low-pass filtering and time-compression between the two levels of distortion with higher levels of distortion yielding prolonged response times. As expected, unaltered words did not differ in their RTs indicating that the RTs were not influenced by their environment (that is, latencies did not vary depending on whether the unaltered target was presented in the severe or the moderate group). A summary of the mean RTs broken up by levels of distortions and perceptual conditions is presented in Table 2.

Additionally, the results yielded a Degree of Manipulation × Perceptual Condition interaction ($F(2, 52) = 13.62$, $p<.0001$). Pairwise comparisons showed significant differences among all auditory conditions except for 1. severe low-pass filtering vs. moderate time-compression and 2. unaltered words presented in the severe vs. moderate group.

Taking the results of the adult subjects together, the following picture emerged:

- Perceptual degradation affected accuracy: Time-compression did not lower the accuracy (compared to unaltered targets), while low-pass filtering did. This effect was independent of the degree of manipulation used. Overall, however, adults had little difficulty correctly interpreting low-pass filtered targets in either degradation level.
- Perceptual degradation negatively affected the subject’s reaction time: Time-compression yielded prolonged response times (measured from the end of the word), followed by low-pass filtering, with the fastest RTs observed for unaltered speech. Although results for accuracy suggest that filtering is harder than compression even for adults, the RT effects suggest that compression has a greater effect on reaction times. Hence we chose the 50% level of compression and the 1.5 kHz level of filtering for the infant study presented next.

EXPERIMENT 2

Method

Participants

Children were recruited through bulk mailing, brochures, advertisements in local parent magazines and visits to postnatal information classes. All of the infants who came to the laboratory were full-term, in good health, with neither pre- nor postnatal complications nor a history of hearing disorders. Our final sample was composed of 95 children ranging from 12 to 31 months of age, 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 3.

An additional 24 infants were tested but not included in the analyses for the following reasons: failure to complete the task (n=11), missing parental language inventories (n=6), experimenter error or equipment failure (n=3), the child’s eye movements were too difficult to track or interference by parents during testing (n=2), fuzziness (n=1) and failure to meet the testing criteria (n=1).

Stimuli

The visual stimuli were the same as those used for the adults. The auditory stimuli were identical to those used in the adult study with the moderate degradation level (= 50% time compression, 1.5kHz low-pass filtering). Additionally, all the target words, their auditory-visual pairings, the sequence of trials, as well as the total number of trials stayed the same.

Apparatus and Procedure

Each child was tested in a sound-proof room. During testing, the child was seated centrally on the parent’s lap, 80 cm in front of a pair of 30cm computers placed about 44cm 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 audio-visual 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, a consent form was signed and the MacArthur Communicative Developmental Inventory was collected (Words and Gestures from 12 to 16 months; Words and Sentences from 17 to 31 months; Fenson et al. 1993). When the child was at ease, both the parent and the child were led into the testing room. The parent was seated in a chair and equipped with opaque dark glasses (so they could not see the target pictures) and headphones playing music selected by the parent. These procedures were adopted to insure that parents could not consciously or unconsciously cue the child regarding the location of the picture corresponding to the auditory name. After verbally encouraging the child to look at a red flashing light located above the monitors, testing was started. Each child was exposed to a total of 48 trials (24 with unaltered targets, 24 with perceptually degraded targets—12 low-pass filtered, 12 time compressed). Successive trials were advanced by the experimenter in the adjacent room only after she determined that the child was fixated on a location midway between the monitors. The experiment lasted on average around 8 to 10 minutes.

Scoring

Because each trial is a complex temporal event yielding different information at different points in time, several measures were applied in analyzing the data. We used a hybrid approach to scoring, combining the methods of Schafer and Plunkett (1998) with those of Fernald et al. (1998, 2001) and Swingley et al. (1998, 1999, 2000). Building on their techniques, we also developed some new measures to look at timing-based “best performance” in an empirically defined “peak” that can vary in its magnitude and timing over children and conditions. This departure from the predefined time windows that are typically used in preferential looking studies was important for the design of the present study, because our perceptual manipulations (unaltered, compressed, and filtered speech) create changes in the point at which a word can be identified. Hence any timing measure based on pre-defined windows would favor one condition over another *a priori*.

Each recording was coded off-line using a button-press apparatus that enabled the experimenter to track the child’s looking times to the right and left pictures. Note that this off-line scoring method has the advantage of reducing experimenter-introduced variability, but it still includes the scorer’s latency, which adds absolute time. Each child was scored a total of 4-8 times—twice for each side separately with one (= 4 scoring runs) or two scorers (=8 scoring runs). Scoring was done by highly experienced scorers who were blind to the position and condition of the target on each trial. The infant’s looking time for each target and distractor was derived from a simple average of the scoring runs for each side respectively (for details see Schafer and Plunkett 1998, Schafer 1998). Intra- and/or inter-scorer relia-bility, assessed for each individual subject using *Correl* was 95% and higher. In order to capture the temporal dynamics of the child’s looking behavior, each scored trial was then cut into 25 ms time slices across its entire length by using the babydat.9 program (developed by Hamilton and modified for our purposes by Yu). In a complex coding scheme (developed by Weir, Zangl and Klarman), each look was coded for target looking time, distractor looking time, away-time (= the time the child was not on task), “peep”-time (= a brief look away from either the target or distractor with a subsequent continuation of the previous look) and shifting time (= the time needed to change from a target to a distractor or vice versa).

Following the Fernald/Swingley (1998, 1999, 2000) method, the trials were divided into three categories, based on where the child was looking at word onset: T-trials (in which children were already looking at the Target picture at word onset), D-trials (in which children were looking at the Distractor picture at word onset) and A-trials (= Away from the task). It is evident that D- and T-trials differ in their cognitive demands: While T-trials require the infants to remain engaged with the picture at which they are already looking, D-trials require them to disengage and shift to the other picture. In our data, D-trials composed a total of 45 % of all trials, followed by T-trials at 41% and A-trials at 14%. A-trials were excluded from all analyses. Accuracy analyses were carried out both by combining and by separating D- and T-trials. Separately analyzing D- and T-trials provided the opportunity to investigate how and to what extent each trial condition was differentially sensitive to perceptual degradation, and also whether they were equally sensitive to developmental changes related to age and/or vocabulary size.

Based on these scoring procedures, four dependent variables were derived: *Epoch Accuracy* (in pre-determined time windows), *Highest Peak Accuracy* or *Peak Amplitude* (a measure of “best” or “peak performance” at variable points in time), *Peak Latency* (a complementary measure of reaction time based on the temporal location of maximum performance or “peak”), and *Peak Duration* (an indicator of the child’s amount of engagement in the speech signal around the time of maximal performance).

Epoch Accuracy

Following earlier studies of infant looking behavior, accuracy was estimated within three pre-defined 1-second time windows or ‘epochs’ starting at target word onset (= Epoch accuracy or EA). Within each epoch, EA refers to the proportion of the time the child spent looking at the target divided by the total time spent looking at both pictures (Target and Distractor). These statistics were calculated first for D- and T-Trials together, and then for D- and T-trials considered separately.

Highest Peak Accuracy (Peak Amplitude)

Highest peak amplitude measures the child's highest level of target recognition at a time point that varies with his/her own looking efficiency, within each perceptual condition. This "peak" or point of maximal performance is defined as the 25-millisecond time slice with the largest number of trials in the correct (in Ds) or incorrect (in Ts) direction. This absolute peak could occur at any time point between 625 ms and 3000 ms. Peaks prior to 625 ms were excluded in order to eliminate "staring" that was independent of the target words (reflecting visual preference and/or random behavior). The 625-ms cut-off was an "educated guess" that included a minimum latency of the child to initiate a shift and a minimum latency of the scorer to press the button box as a reaction to the child's looking behavior (Hood & Atkinson, 1993, Haith et al. 1993). In Ds the peak amplitude denotes maximum accuracy (= a higher amplitude corresponds to a better match), while in Ts the peak amplitude denotes maximally incorrect looking behavior. The "peaks and valleys" yielded by these measures are modeled after metrics that have been used successfully for many years in the electrophysiological literature, to quantify event-related brain potentials. Peak amplitude was calculated separately for each condition (C, F, U) and trial type (D-, T-trials), for each child.

Peak Latency

Reaction times were derived by measuring the amount of time each child needed to reach the highest peak (= peak latency). Peak latency (like Peak Amplitude) was calculated separately for each condition (C, F, U) and trial type (D-, T-trials), for each child. In D-trials, peak latency denotes the amount of time in ms needed to reach the highest point in accurately shifting from D to T. In T-trials it denotes the amount of time in ms needed to reach the highest point in inaccurately shifting from T to D. If peak latency is influenced by the acoustic quality of the target word and/or the child's age/vocabulary variations in timing are expected.

Peak Duration

Stimulus engagement was measured by calculating peak duration times in D-trials. Peak duration in D-trials is defined as the total time spent at maximum accuracy. In addition to peak duration, 'pure' target looking time (= 100% on target) was calculated for T-trials only – once from the onset to the first indication of change, and again from word onset to 3000 ms. Since 'pure' target looking times were very rare in D-trials, no comparable measurement could be taken.

Results

Analyses of variance were conducted twice for each dependent variable—once with participants grouped by age and another with participants grouped by expressive vocabulary size. In both cases children were assigned to one of four groups. For the analysis by age, the four groups were: 12-14 months (n=22; mean age: 12.6), 15-18 months (n=25; mean age: 16.4), 19-23 months (n=30; mean age: 20.6) and 24-31 (n=18; months (mean age: 25.7). For the analysis by Vocabulary the four groups were: 0-20 words (n=27; mean words: 6.9), 21-99 words (n=31; mean words: 52.9), 100-300 words (n=21; mean words: 183.2) and >300 words (n=16; mean words: 462.4). (Note that vocabulary was based on number of words *produced* on the parental report measure, and not on number of words comprehended). 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 et al. 1995, 1999). The split into four vocabulary groups was based on previous research using the MacArthur Communicative Developmental Inventory as well as careful preliminary analyses of the MacArthur data for our sample. Both suggested break-points in the developmental curves for vocabulary production at the >20, >100, and >300 word level. Within the 0-20 word level, words are integrated in a piece-meal fashion; i.e. the child's vocabulary expands at a slow rate. The rate of growth catches up in the next group with a faster integration of new words typically from around 50 words onwards. Between 100 and 300 words word combinations get off the ground together with a more rapid integration of new words. Above 300 words the development of grammar is well under way; the children's language has increased not only in mean length of utterance (MLU) but also in the use of grammatical morphemes. In the same vein, age groupings reflect the average windows within which changes in the speed and composition of vocabulary have been observed in previous studies.

Accuracy

The central question to be addressed with both accuracy measures was whether children's looking behavior differed between the three auditory conditions. If performance is vulnerable to perceptual degradation, children's accuracy should be inferior in time-compressed and low-pass filtered words. Differences between low-pass filtering and time-compression would reflect differential sensitivity to speeded versus spectrally poor signals.

Epoch Accuracy (=EA)

We started off with an analysis of Epoch Accuracy (proportion of trials looking at the target) with **D- and T-trials combined**. The respective 4 (age or vocabulary) × 3 (perceptual condition: C, F, U) × 3 (epoch: 1st, 2nd, 3rd second) mixed analyses of variance yielded significant main effects for both of the between-subject factors (age and vocabulary). The main effect of Vocabulary (F(3, 94) = 6.78, p<.0004) was a little stronger than that of Age (F(3, 94) = 4.54, p<.005). In both cases, accuracy increased across developmental levels as shown in Figures 2A and 2B (*note that target fixations were measured across all conditions).

Main effects of Perceptual Condition and of Epoch were also significant in both analyses, all at the p<.0001 level (Vocabulary - Perceptual Condition: F(3, 94) = 33.54, Epoch: F(3, 94) = 74.75; Age – Perceptual Condition: (3,94)=35.59, Epoch: F(3, 94) = 67.87). The effect of Perceptual Condition reflected significantly lower accuracy for filtered (F) stimuli staying at chance (50%) followed by compressed (C, 56%) and unaltered (U, 59.6%) trials. There were significant differences between each of the conditions as indicated by planned comparisons (C vs. U: t(94) = 10.67, p<.001; C vs. F: t(94) = 26.04, p<.0001; F vs. U: t(94) = 70.05, p<.0001).

The effect of Epoch reflected an increase in proportion of time looking at the target, which however was not linear. The children's correct target looks sharply increased from the 1st to the 2nd time window (with planned comparisons in both groupings at p<.0001), and then dropped significantly between the 2nd and 3rd time window (as demonstrated in Figure 3). This clearly indicates that target recognition really "set in" during the 2nd epoch (and despite a decline stayed above chance in the 3rd epoch).

In addition to the main effects, the analysis of age yielded two-way interactions of Perceptual Condition × Age, of Epoch × Age and of Perceptual Condition × Epoch as well as a three-way interaction of Perceptual Condition × Epoch × Age (for the statistics of the interactions see Table 4). The interactions are plotted in Figures 4A to 4C.

A closer inspection of the Perceptual Condition × Age interaction revealed that it occurred because of the stagnation of filtered stimuli across age; in contrast, compressed and unaltered words reliably increased with age. The two-way interaction of Epoch × Age was caused by differences between the 1st vs. the 2nd and 3rd Epoch, a difference that was more marked for older children. The Perceptual Condition × Epoch interaction occurred because the three perceptual conditions diverged across the time window: compressed and unaltered words increased over time while filtered remained continuously low across all 3 seconds. Finally, the three-way interaction confirms the trend just explained: compressed and filtered words increased in accuracy over time from 12 to 24 months; filtered words were "frozen" over time and age (for the statistics see Table 4).

Using vocabulary as a grouping factor yielded the same interactions as for age, with one noteworthy difference: the two-way interaction of Perceptual Condition × Age did not surface as an interaction between Perceptual Condition × Vocabulary. Specifically, Vocabulary revealed a developmental change in the filtered condition that was not detected in the analysis by Age: filtered stimuli "joined" the other conditions, demonstrating increased target accuracy across epochs in children with higher vocabularies, even for the low-pass filtered words. This is a particularly useful finding, because it means that growth in vocabulary is associated with the emerging ability to extract enough information from a filtered stimulus to achieve accurate word recognition. A summary of the interactions for the age and vocabulary groupings is presented in Table 4 below.

Since target recognition was most clearly demonstrated in the 2nd time window, some further post-hoc analyses were carried out within this epoch only. As evident in Figure 5 there was a large effect of Perceptual Condition (p<.0001) with significantly higher accuracy scores for unaltered (65%) and compressed (63%) than filtered (52%) stimuli. Simple effects of the between-subjects variables were also computed within this epoch, once for age and again for vocabulary. While the function of age revealed significant effects for compressed (F(3, 94) = 7.89, p<.0007) and unaltered (F(3, 94) = 17.12, p<.001) but not filtered stimuli, the function of vocabulary showed significant developmental change in all three conditions. *In other words, a developmental increase for filtered words was only seen when children are grouped by vocabulary.*

Based on prior work by Fernald, Swingley and colleagues, we know that D- and T-trials often yield dramatically different results. To determine whether the above results for epoch accuracy differ over these two trials types, the analyses were repeated for D- and T-trials separately. For **D-trials only,** results replicated the above findings for D+T together. This includes the finding that performance in the filtered condition is better predicted by vocabulary and than age. This finding for D-trials alone provides useful information about the robustness of word recognition in this age range (at least for the more sophisticated children): even when starting off at the incorrect picture, children managed to fixate on the correct visual image in filtered conditions. Figures 6A, 6B and 6C compare the developmental functions in Epoch Accuracy (measured within the 2nd epoch) for age and vocabulary, respectively. Post-hoc explorations of effects within developmental levels showed that children below 15 months or below the 20-word level failed to establish correct matches in any of the auditory conditions. From 15 months onwards or above the 20-word level there was clear evidence of target identification for both compressed and unaltered words. In both conditions we observed two significant shifts – one from below to above chance level from 0-20 words to 21-99 words and the other one from 100-300 words to > 300 words. Between 99 and 300 words development had reached a plateau. Correct identification of filtered words, on the other hand, did not start until children had a vocabulary of 100 words; it stayed at chance in the age grouping. Note the similarity of curves for compressed and unaltered words (both for age and vocabulary) and the performance "gap" that opened up in filtered words.

In order to complete the picture, we computed separate Epoch-Accuracy analyses for **T-trials only**. As mentioned earlier T-trials are cognitively less challenging since the child is already at the correct match when the target word comes on. Thus, they provide information about *incorrect* shifting behavior by the young listener. While analyses of both Perceptual Condition and Epoch were highly significant for both groupings (p<.0001), neither Age nor Vocabulary showed a main effect. This suggests that T-trials are less developmentally sensitive than D-trials (at least in Epoch Accuracy), more or less stagnating across age or vocabulary size (collapsed over all conditions and all epochs). Interestingly even when starting off at the correct picture, compressed and unaltered words are 'yoked' together over epochs and differ from filtered words: there was a sharper increase in incorrect looks in filtered vs. compressed and unaltered words. Furthermore, correct target looks decreased over epochs with the steepest decrease again between the 1st and 2nd time window (from 20% to 37% incorrect looks collapsed over all conditions). In addition to the main effects there was a highly significant two-way interaction of Perceptual Condition × Epoch and a three-way interaction of Perceptual Condition × Epoch × Age/Vocabulary (for the statistics see Table 5). The two-way interaction was caused by differences between the 1st and 2nd/3rd Epoch, which showed a steep increase in incorrect looking behavior (that is, looking away from the target) compared to correct fixations in the early time window. Interestingly, filtered words in the highest age/vocabulary group showed a different looking pattern in the 3rd epoch. While children kept fixating the incorrect picture in both the compressed and unaltered condition, they shifted back to the correct picture in the filtered condition. What this means is that, in the filtered speech condition, these children first looked at the target, then did a quick check of the distractor, and then switched back to the correct target and remained there.

Highest Peak Accuracy (Peak Amplitude)

The above analyses of epoch accuracy establish that children's looking behavior is indeed affected by the target words, and that this behavior changes over time and levels of development, in patterns that differ across perceptual conditions. These results were based on the kinds of fixed time windows that have been used in previous studies, providing continuity to the larger literature on preferential looking. Our second measure of accuracy, Highest Peak Amplitude, serves two further purposes: (1) it provides an additional measure of accuracy that is tied to an empirically defined metric that can vary over the markedly different timing conditions utilized in this experiment, and (2) it permits us to establish a marker for peak performance that can be used to derive further information about reaction time (Peak Latency) and stimulus engagement (Peak Duration). Because the definition of "peak" is quite different for D- vs. T-trials, analyses will be conducted separately.

D-Trials

Using Age as a grouping variable, a 4 (age) × 3 (perceptual condition) mixed ANOVA demonstrated main effects of Age (F(3,94)=3.78, p<.01) and of Perceptual condition (F(2,94)=9.98, p<.0001). The effect of Age reflected an increase in peak amplitude over age. The developmental curve showed a non-linear increase with two shifts: one from the 12-month to the 15-month and another from the 19- to the 24-month groups; between 15 and 19 months performance had reached an temporary standstill as seen in Figure 7.

The developmental shifts shown in Figure 7 resemble the ones obtained for Epoch Accuracy. In post-hoc analyses within each perceptual condition, the increase with age only reached significance for unaltered targets (F(3, 94) =10.48, p<.0001). The effect of Perceptual Condition reflected lower peak amplitudes for filtered (69%) compared to compressed (80%) and unaltered (77%) stimuli but (as shown by planned comparisons) no difference between compressed and unaltered stimuli. In post-hoc analyses looking at each age group separately, the main effect of perceptual condition reached significance in the 15- and 24-month-olds only. In both groups the perceptual effect was reflected by lower peak amplitudes for filtered targets.

A separate analysis by Vocabulary grouping produced not only stronger but also developmentally more sensitive results. Main effects were evident for both Vocabulary (F(3, 94) = 8.36, p<.0001) and Perceptual Condition (F(2, 94) = 8.93, p<.0003). Children with higher vocabularies had higher target-looking scores. In contrast to age, the developmental curve (plotted in Figure 8) was linear with significant shifts between each vocabulary level .

The effect of Perceptual Condition corresponded to the one achieved for Age. Children's peak amplitudes were significantly lower in filtered than in compressed and unaltered conditions. A separate analysis within each condition showed effects of vocabulary size in compressed (F(3, 94) = 4,58, p<.005) and unaltered (F(3, 94) = 9.58, p<.0001) words, and a trend towards significance in filtered stimuli (F(3, 94) = 2.57, p<.06). Children with higher vocabularies had higher peak amplitudes than children with low vocabularies. This suggests that they were not only more correct overall (as evident in the epoch analyses), but also reached higher peaks. To determine whether age and vocabulary were contributing separate variance, the analysis by age level was repeated controlling for vocabulary size (treated as a continuous variable). With this control, the main effect of age failed to reach significance. The analysis by vocabulary level was then repeated controlling for age (with age treated as a continuous variable). In this case, a significant effect of vocabulary level remained, at even higher levels of significance (F(3, 94) = 3.24, p<.03*). In other words, we can conclude that developmental variance in looking accuracy is driven by vocabulary level independent of the variance contributed by age alone.*

<u>T-trials</u>

Contrary to D-trials, neither Age nor Vocabulary revealed a main effect in T-trials. However, T-trials differed in relation to signal quality. Perceptual Condition was significant at the p<.0001 level, with larger incorrect amplitudes in filtered (69%) **and** compressed (64%) targets, compared with unaltered (53%) targets. It seems that in T-trials both the compressed and filtered words yielded more error. This is the only analysis so far that suggests a separation between "stressed" (both compressed and filtered) vs. "normal" processing.

Peak latency

As indicated earlier, peak latency is a measure of response time that reflects the time required to reach peak performance. Like peak amplitude, peak latency can vary over children and conditions. Because the definition and direction of peak location differs for D- and T-trials, these events were (again) analyzed separately.

<u>D-trials</u>

Significant main effects were observed for both Age (F(3, 94)=5.002, p<.003) and Vocabulary (F(3, 94)=7.38, p<.0002). Once again, when the age analysis was repeated controlling for vocabulary size, the effect disappeared, but when the vocabulary analysis was repeated controlling for age, the effect remained significant (F(3, 94) = 5.66, p<.0001). Interestingly, RTs were consistently slow until the 24months/>300 word-point and then sharply dropped as evident in Figure 9. While the oldest/most sophisticated group "peaked" at 1325 ms (for age) and at 1241 ms (for vocabulary) after target word onset, all the other groups were about 400-500 ms slower consistenly "peaking" around 1700 ms (note that the peak latencies were averaged across conditions). This sharp drop is quite different from the more gradual changes observed over age and vocabulary level for most of the accuracy measures, and it suggests a rather dramatic reorganization in the efficiency of word recognition around this window of development. We will say more about this issue in the final discussion.

A significant main effect of Perceptual Condition was found in both groupings (at p<.02, F(3, 94) = 4.17), reflecting faster response latency with compressed stimuli compared to unaltered but not filtered stimuli. Decreased latencies for compressed words may be due to the fact that complete information about the identity of the word arrives twice as fast for compressed stimuli than the other two words types—a fact that children in this age range seem to be able to exploit. Separate analyses for each condition for both age and vocabulary showed decreased latencies with increased age/ vocabulary for compressed targets (age: F(3, 94) = 3.10, p.<.03, vocabulary: F(3, 94) = 4.82, p<.004) and unaltered targets (age: F(3, 94) = 5.27, p<.002, vocabulary: F(3, 94) = 8.19, p<.0001). There was no significant developmental change in response latencies for filtered targets in either grouping.

<u>T-trials</u>

Contrary to D-trials, peak latency in T-trials measures the amount of time needed to reach the highest peak in *incorrect* shifts (i.e. maximum time off target). Neither Age nor Vocabulary produced a significant effect, suggesting that peak latency in T-trials is not a developmentally sensitive measure. However, peak latencies for T-trials did vary in relation to Perceptual Condition (F(3, 94) = 3.97, p<.02), reflecting a faster shift away from the target for compressed words. This last result may help us to understand why peak latency is developmentally insensitive for T-trials even though it proved (above) to be very sensitive to developmental variables on D-trials. For developmentally less sophisticated children, these shifts away from the right picture may simply reflect a return to baseline (randomly looking back and forth from one picture to another) when the sound stimulus comes to an end (something that occurs earlier for compressed trials). For more sophisticated children, the same behavior may have a different meaning, with children conducting an earlier “compare and check” for compressed trials (as if to say “Yes, that’s the picture with the dog”). Whether or not this speculation is correct, the fact that reaction time findings differ so markedly for D- and T-trials underscores the value of distinguishing between these two kinds of looking events, as Fernald, Swingley and colleagues have noted in their work.

Peak duration

This measure was designed to assess the physical “attractiveness” of the speech signal and its matching picture. It may also assess the amount of time that children require to satisfy themselves that a match has been made. Children may be more interested in stimuli that are more complex or unusual, and/or they may stay engaged for a longer period because they are still ‘solving the problem’. Since perceptual degradation enhances the complexity of the signal, higher looking times (= increased ‘stickiness’) would be expected in acoustically modified targets compared to unaltered ones. Peak duration was determined by calculating the time from the onset of the highest peak to its offset, and is restricted (based on the definition of peak amplitude) to the period between 625 and 3000 ms. The total amount of time is referred to as peak duration time, and is conducted for D trials only. In addition, ‘pure’ target looking times (= 100% at the target) were calculated in T-trials. Since pure target looking times were rare in D-trials no comparable measurements could be taken.

Peak Duration in D-trials

Separate ANOVAs—one for age, one for vocabulary—both revealed significant main effects of Perceptual Condition at the p<.0001 level, as can be seen in Figure 10. However, there were no main effects of either Age or Vocabulary. Once the peak was reached, children stayed significantly longer at compressed than at filtered or unaltered targets across all age and vocabulary levels. The mean looking time for compressed targets was 597 ms vs. 465 ms for filtered and 330 ms for unaltered words. Planned comparisons indicated significant differences between each of the conditions (C vs. F: t(94) = 4.28, p<.04, F vs. N: t(94) = 4.00, p<.05 and C vs. N: t(94) = 16.53, p<.0001). This suggests that perceptual degradation does enhance stimulus engagement. Simply put, children stare harder and longer when stimuli are more difficult. Combining the information from peak duration with the above results for peak latency, we also have a clearer picture of the results for compressed trials: children may move quickly toward the target on compressed trials (because the information is available sooner) but they also tend to “stick around” longer, as if they are taking some time to be sure that the picture corresponds to what they just heard.

‘Pure’ Target looking times in T-trials

In agreement with the results above, perceptual degradation also influenced target “stickiness” in ‘pure’ target looking times for T trials. The main effect for Perceptual Condition was significant at the p<.0001 level with target-looking times varying across the auditory conditions as evident in Figure 11.

The data suggest that perceptual degradation influenced the children’s looking behavior right from target word onset, as well as later on in the trial, with significantly higher engagement times with compressed (939 ms) and filtered (766 ms) targets than with unaltered (472 ms) ones (=total target looking times). When looking at the total amount of ‘pure’ target looking time, both factors were significant, with Vocabulary stronger (F(3, 94) = 4.65, p<.005) than Age (F(3, 94) = 3.82, p<.01). In both groupings, lowest/youngest and highest/oldest groups drifted off faster than the groups in between. If ‘stickiness’ or failure to disengage reflects the amount of time that children require to be sure they are right, then we may speculate that the least advanced children disengage quickly because they simply do not get the point at all, while the most sophisticated children disengage quickly because they have solved the problem quickly and efficiently, to their own satisfaction, and are anxious to get on with things.

Summary & Discussion

Processing speech under adverse conditions has been investigated in adults and older children; how the developing infant reacts to less “perfect” speech is unclear. Generally, the infant may be exposed to a variety of processing climates ranging from optimal “hyperspeech” (= infant-directed speech; Fernald 2000) to more taxing speech which can refer to speech that either varies in speed or in clarity (e.g. wirh background noise). Although previous studies have shown that infant-directed speech is very common in interactions with infants (Cooper& Aslin, 1990, 1994; Cooper 1993, Fernald 1985, Fernald et al. 1989), children also face variations in speed of different talkers or the "newness"(given vs. new contrast) of the information (Fisher & Tokura, 1995). Using stimuli differing in their acoustic qualities ranging from prototypical infant-directed (=ID), to time-compressed (showing ID-pitch characteristics but reduced duration), to spectrally degraded speech (showing ID-pitch characteristics but reduced segmental information), the present study enabled us to assess the impact of different processing climates on early word processing and to explore the robustness of the infant's system.

Within this framework, the present study had three goals: (1) to compare the effects of perceptual degradation in experienced adult listeners with the effects obtained in infants who are in the first stages of word learning, using two different forms of perceptual degradation—low-pass filtering and temporal compression; (2) to investigate the development of word recognition under all three conditions (unaltered, filtered, compressed) across the period from 12 - 31 months, comparing effects of age with effects that are obtained when children are grouped by vocabulary level; (3) to explore the utility of "dynamic" measures of looking behavior for the assessment of developmental changes in real-time word recognition during the early stages of language development. Methodological innovations include variations over trial type (based on where the child was looking at word onset, following Fernald, Swingley and colleagues) and some novel measures of looking dynamics (Peak Amplitude, Peak Latency and Peak Duration), inspired by measures used in the electrophysiological literature. Let us briefly review results of Experiments 1 and 2 in the light of these three goals.

(1) Comparing Children and Adults

Experiment 1 with adults served two purposes. First, it permitted us to select appropriate degrees of distortion for use with infants. Second (and more interesting for our purposes here), it permitted us to compare the emergence of word recognition in the early stages of development with performance by experienced listeners (i.e. the "optimal system"). We found that perceptual degradation did affect word processing both in adults as well as in infants. The effects, however, differed in their magnitude and characteristics.

Low-pass filtering had devastating effects on word recognition in infants, but the same infants had little or no difficulty dealing with 50% time compression (at least of infant-directed speech). In contrast, adults had little difficulty with either form of perceptual degradation. They were able to identify low-pass filtered targets at high levels of accuracy even in the more severe distortion level. The devastating effects of reduced spectral information for children is in line with the studies of older children by Eisenberg (2000) and Dorman (1998b), who observed long-term developmental learning effects under conditions of spectral degradations. 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 just how much degradation is required to set off these effects. Using word-initial mispronunciations, Swingley & Aslin (2000) documented significant reductions in looking speed and accuracy, but those effects were much milder than ours.

In contrast with the sharp differences in effects of filtering on infants vs. adults, adult-child differences in the effects of temporal compression were quite subtle. When RTs are measured from the end of the word (as they were in our analyses of adults), we could see that adults are able to identify unaltered words before all of information had arrived; for compressed targets (despite the advantage of 'faster arrival'), compression did exact a small cost, and recognition times tended to occur after the end of the word. For the infants in this study, effects of compression suggest that the advantages of early word arrival outweigh the disadvantages: by at least 15 months of age, they are able to identify compressed words just as accurately as unaltered words, and they are able to take advantage of the early arrival of lexical information for compressed trials, shifting (peak latency in D-trials) as soon as the relevant information is available. The only evidence in this study for a "compression cost" in infants came in the measure of peak duration, which suggested that children tend to stare longer after they reach peak accuracy for compressed words, as if they were 'double checking' to make sure that they made a correct match (discussed further below). To explain the sharp contrast between filtering and compression in the present study, we tentatively suggest that the fragile lexical representations that underlie word comprehension in infants require a greater amount of "bottom up" perceptual information (reduced with low-pass filtering); however, if enough information is available in the signal, they can withstand (within the limits manipulated here) speeded presentation of that information (see Utman & Bates for a further discussion of this issue).

These results for temporally compressed words are also relevant to a large body of work on the nature and purpose of Infant-Directed Speech (ID-Speech). By 15 months of age, speaking rates up to twice as fast as normal (where normal = normal for infant-directed speech) provide no problems in comprehension. We may therefore speculate that increased duration is not necessary for word comprehension -- as long as pitch characteristics and signal clarity are still preserved. From the point of view of a parent or other caretaker, these results suggest that it is not necessary to strategically adjust the length of words to the child's age/vocabulary level. Segmental clarity seems to be more important than duration for the benefit of the infant.

(2) Vocabulary vs. Age as Predictors of On-Line Word Recognition

With regard to the second goal, we found significant effects of both age and vocabulary on infant word recognition. However, vocabulary proved to be a more sensitive developmental indicator than age for both accuracy and response times, and remained significant when effects of age were controlled. More specifically, grouping by age did not reveal any developmental changes from 12 to 31 months in filtered words, but grouping by vocabulary did reveal developmental change for these especially devastating stimuli. It is important to keep in mind that our measure of vocabulary was based on the number of words that children are reported to produce. The fact that this variable proved to be a powerful predictor of listening behavior suggests to us that there are major changes in the efficiency of lexical processing around the so-called "vocabulary burst", which are reflected in on-line comprehension as well as expressive vocabulary size (see also Reznick & Goldfield, 1992)

Swingley & Aslin report that effects of word-initial mispronunciations were not related to age or to spoken vocabulary size in their study. However, other studies have documented a relationship of vocabulary size and children's performance patterns in this age range (Fernald et al. 2001, Mills et al 1997, Mills 1999, Bates & Goodman 1997). A relationship between vocabulary size and spoken word recognition has recently also been reported in older children (3 to 8 years of age; Munson in press). Using two spoken word recognition tasks – gated words (where the final stop in a CVC-combination was deleted) and noise-center words (where the initial vowel was replaced by broad-band noise)—Munson could demonstrate that receptive and productive vocabulary were reliable predictors of a significant proportion of variance in the word recognition scores. Interestingly, vocabulary was a better predictor than other measures such as pre-literacy skills, phonological awareness and articulation accuracy. These results (although for older children) are in line with our findings here.

(3) Dynamics of word recognition: measurement comparisons

Our third goal was to compare different measures of real-time word recognition in this looking paradigm, to see if these measures give us distinct but complementary information about developmental change between 12-31 months of age. In addition to a more traditional measure of looking accuracy within pre-defined time windows or epochs, we identified for each child, for each condition, an empirically defined point of "peak performance." These peaks were used to derive temporally sensitive measures of accuracy (peak amplitude), reaction time (peak latency) and stimulus engagement (peak duration). All measures were applied separately to distractor trials (D-trials) and target trials (T-trials), building on prior work by Fernald, Swingley and colleagues who have shown that these two distinct looking events yield very different perspectives on the development of word recognition. In the present study, this combination of looking measures and separate analyses by trial type provided complementary information on the emergence and development of listening and looking skills during the first stages of word learning.

First let us consider the differential results obtained for trial types. Analyses of distractor trials (D trials) revealed developmental changes in target fixation with both age and vocabulary. In contrast, there was little evidence for developmental change on target trials (T-trials) at any point across the period between 12 and 31 months. This suggests that situations in which the child starts off at the wrong picture and has to initiate a shift to the other picture to perform a correct match were developmentally more informative than situations in which the child starts off at the right match and 'only' has to remain engaged in what she was already doing. However, performance in both trial types was influenced by the respective perceptual conditions, which indicates that—independent of the child's locus at word onset—subsequent looking behavior was modulated by the acoustic-phonetic shape of the signal. On D-trials, where children have to shift from the "wrong" to the "right" picture, accuracy was equally high for compressed and unaltered stimuli; in contrast, only the more sophisticated children (with vocabularies above 100 words) showed accurate shifts for filtered stimuli. On T-trials, where children have to stay where they are in order to obtain high accuracy scores, we see a greater tendency to drift back toward the wrong pictures on both compressed and filtered trials. At the very least, these results underscore the utility of separate analyses for looking events that require “moving on” vs. events that require “hanging in”.

Because our measures of accuracy, latency and stimulus engagement yield somewhat different perspectives on developmental change from 12-31 months, results are summarized separately.

Accuracy

Applying two accuracy measures provided the opportunity to assess whether or not perceptual degradation affected performance consistently across different operational definitions of “getting it right”. The advantage of a combination of the two measures lies in the fact that accuracy could be assessed in two qualitatively different ways: one using pre-defined time windows each stretching over a 1 sec epoch (the traditional approach in studies using this paradigm); and the other exploiting the child’s own looking efficiency and picking up the highest point in performance. Independent of the calibration used, the basic results were the same: (1) Perceptual degradation did affect performance in both measures, and (2) children with higher vocabularies/age levels identified targets more reliably than did children with lower vocabularies/age levels. Most important for our purposes here, developmental changes in accuracy differed as a function of perceptual conditions. For the youngest (12-15 months) and least sophisticated children (<20 words), there was little evidence of word recognition in any of the three perceptual conditions. By 15 months (and >20 words), accuracy increased over epochs on both unaltered and compressed trials. For filtered stimuli, there was no evidence for a developmental increase in accuracy when children were grouped by age, and when they were grouped by vocabulary size, evidence for accurate recognition of filtered stimuli was only evident in the more linguistically sophisticated groups (>100 words).

The accuracy measures also yielded some interesting insights into patterns of growth. When children were grouped by age, we found a non-linear profile with two significant points of increase surrounding a long plateau (between 15-24 months). When the same children were grouped by vocabulary, a more linear pattern of change emerged across all levels of vocabulary size. Thus, depending on the respective perceptual condition, correct target identification revealed different starting points as well as different growth patterns (evident in both accuracy measures). Because correct identification of filtered words started at a time when many of the children’s vocabularies were spurting, we propose that the vocabulary spurt (at least in part) is related to an improved ability to handle spectrally imperfect speech signals. It may well be that children with larger vocabularies have more robust and well-defined lexical representations, which in turn are more resistant to demanding processing climates. The vocabulary spurt has also been associated with more general cognitive changes such as perceptual and attentional advances which may in turn help the child to cope with more diverse and also more complex processing conditions (Bloom 1993, Roberts 1998). In short, the ability to correctly identify unaltered words may depend on a complex interplay between language-specific and more general, cognitive factors.

Speed

Our measure of peak latency (for D-trials only) proved to be quite sensitive to both age/vocabulary and perceptual conditions. With regard to developmental change, the most dramatic change was visible in the >300 word group, who showed a tremendous increase in processing speed -- a 500 millisecond drop in peak latency, collapsed over all conditions. This non-linear drop in reaction times is quite different from the gradual increase observed in accuracy when children are grouped by the same vocabulary levels. This non-linear change in processing efficiency directly mirrors the non-linear “spurt” or acceleration in expressive vocabulary that also occurs between 200-400 words in most children, providing further evidence for a synergistic relationship between speed of word learning and efficiency in word comprehension. With regard to perceptual conditions, peak latencies were faster for compressed targets, suggesting that infants are able to exploit the early arrival of information that these stimuli provide.

Peak Duration

Peak duration was used as a measure of the child’s degree of engagement in the stimulus. As we expected, this measure was also affected by perceptual degradation. The visual engagement measure provided insights into how long the child was fixated on the visual image when at the correct peak (as indicated by peak duration times in distractor trials or ‘pure’ target times in target trials). How long the child fixated on the picture may have been influenced by several factors: attentiveness to the task, alertness, stimulus complexity, speed of encoding the visual image, and/or integration of visual and auditory cues. If perceptual degradation generally increases the processing demand (and/or alertness and attention of the child), then longer fixation times are expected with increased cognitive complexity. This is exactly what our analyses revealed. Perceptual degradation resulted in a substantial difference in peak target fixation times: children engaged longer in time-compressed and low-pass filtered than in unaltered words. Longer visual fixation might be linked to the effort required in processing perceptually degraded stimuli (i.e., children were 'working harder'). By increasing the stimulus complexity additional attentional resources were recruited, with compressed and filtered words functioning as attentional spotlights (for the relationship between task difficulty and attention see Kahneman 1973, Norman & Bobrow 1975, Pashler 1993, 1998). Although peak duration was not sensitive to either age or vocabulary in distractor trials, effects related to vocabulary size were seen for target trials, but the pattern was non-monotonic. In target trials we observed less engagement in the lowest and highest vocabulary groups. Decreased engagement at the opposite ends may be differentially motivated: In the 0-20 word group target recognition is at chance which might lead to reduced stimulus-related engagement in general. In the >300 word group, target recognition is well developed. Encoding and integration of auditory visual information may require less time and be also less demanding, which might lead to faster disengagement.

In summary, the present study demonstrates that early word processing is strongly affected by signal characteristics. This is also true for word processing in more experienced adults. In the latter case, overall patterns of results were similar to those of children, but they differed in the magnitude of the effects. The results showed differential effects of low-pass filtering versus time-compression in early (as well as in advanced) language processing. In infants, low-pass filtering clearly affected successful processing more profoundly than did time compression. It severely and consistently reduced the children's ability to correctly identify the target picture. Time compression, on the other hand, did not markedly decrease the children's performance. Furthermore, the stronger predictive strength of vocabulary over age in relation to developmental patterns (which were especially pronounced with increased signal complexity) can be seen as an important indication for future studies.

Acknowledgements

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, and to Lin Chang, Ramona Friedman, Bryce Pearson, Leah Shuchter, Alexandra Wydogen and Sheri Yoshimi for assistance in data collection. We are especially grateful to Anne Fernald, Kim Plunkett, Graham Schafer, Alycia Cummings and Amy Perfors for detailed advice at various stages of the project. In addition, we are grateful to Melissa Schweisguth for her support in experimental set up and the collection and analysis of adult data.

REFERENCES

Balaban, M.T., & Waxman, S.R. (1997). Do words facilitate object categorization in 9-month-old infants? *Journal of Experimental Child Psychology* 64 (1), 3-26.

Bates, E., Bretherton, I. & Snyder, L. (1988). From first words to grammar: individual differences and dissociable mechanisms. Cambridge: CUP.

Bates, E. & Goodman, J. (1997). On the inseparability of grammar and the lexicon: Evidence from acquisition, aphasia and real-time processing. In: G. Altmann (Ed.), Special issue on the lexicon, *Language and Cognitive Processes,* 12 (5/6), 507-586.

Bloom, P. (1993). The transition from infancy to language: Acquiring the power of expression. Cambridge: CUP.

Blackwell , A., & Bates, E. (1995). Inducing agrammatic profiles in normals: Evidence for the selective vulnerability of morphology under cognitive resource limitation. *Journal of Cognitive Neuroscience* 7 (2), 228-257.

Blumstein, S.E., Katz, B., Goodglass, H., Shrier, R., & Dworetsky, B. (1985). The effects of slowed speech on auditory comprehension in aphasia. *Brain and Language 24,* 246-265.

Caselli, M.C., Bates, E., Casadio, P., Fenson, J., Fenson, L., Sanderl, L., & Weir, J. (1995). A cross-linguistic study of early lexical development. *Cognitive Development* 10, 159-199.

Caselli, M.C., Casadio, P., & Bates, E. (1999). A comparison of the transition from first words to grammar in English and Italian. *Journal of Child Language* 26, 69-111.

Charles-Luce, J., & Luce, P.A. (1990). Similarity neigbourhoods of words in young children's lexicons. *Journal of Child Language* 17, 205-215.

Charles-Luce, J., & Luce, P.A. (1995). An examination of similarity neighbourhoods in young children's receptive vocabularies. *Journal of Child Language* 22, 727-735.

Connolly, J.F., Stewart, S.H., & Phillips, N.A. (1990). The effects of processing requirements on neurophysiological responses to spoken sentences. *Brain and Language* 39 (2), 302-318.

Cooper, R., Ostroff, W.L., Cooper, J.S. & Ward, C. (2000). The usefulness of looking time as a measure of inants' speech preferences through-out the first year of life. Presented at the *International Society for Infant Studies.* Brighton, England.

Cooper, R. P., & Aslin, R., N. (1994). Developmental differences in infant attention to the spectral properties of infant-directed speech. *Child Development* 65, 1663-1677.

Cooper, R.P, (1993). The effect of prosody on young infants' speech perception. In C. Rovee-Collier & L.P. Lipsitt (Eds.), *Advances in infancy research Vol. 8, 137-167.* Norwood, NJ: Ablex.

Cooper, R.P., & Aslin,R.N. (1990). Preference of infant-directed speech in the first month after birth. *Child Development 61, 1584-1595.*

Devescovi, A., Pizzamiglio, L., Bates, E., Hernandez, A., & Marangolo, P. (1994). Grammatical deficits in patients without agrammatism: Detection of agreement errors by Italian aphasics and controls (Abstract). *Brain and Language* 47 (3), #17.

Dick, F., Bates, E., Wulfeck, B., Utman, J., Dronkers, N., & Gernsbacher, M. (2001). Language deficits, localization and grammar: Evidence for a distributive model of language breakdown in aphasics and normals. *Psychological Review, 108*(4), 759-788**.**

Dorman, M.F., Loizou, P.C., Kirk, K.L., & Svirsky, M. (1998b). Channels, children and the multisyllabic lexical neighbourhood test (MLNT), NIH Neural Prosthesis Workshop, Bethesda, MD, October 1998.

Dupoux, E., & Green, K. (1997). Perceptual adjustment to highly compressed speech: effects of talker and rate changes. *Journal of Experimental Psychology: Human Perception and Performance* 23 (3), 914-927.

Eisenberg, L.S., Shannon, R.V., Schaefer Martinez, A., & Wygonski, J. (2000). Speech recognition with reduced spectral cues as a function of age. *Journal of the Acoustical Society of America 107,* 2704-2710.

Fenson, L., Dale, P., Reznick, J.S., Thal, D., Bates, E., Hartung, J., Pethick, S. & Reilly, J.S. (1993). *Mac Arthur Communicative Developmental Inventories: User's guide and technical manual.* San Diego, CA: Singular Publishing Group.

Fernald, A., Swingley, D. & Pinto, J.P. (in press). When half a word is enough: infants can recognize spoken words using partial phonetic information. *Child Development.*

Fernald, A. (2000). Speech to infants as hyperspeech. Knowledge-driven processes in early word recognition. *Phonetica 57, 242-254.*

Fernald, A., Pinto, J.P., Swingley, D., Weinberg, A. & McRoberts, G.W. (1998). Rapid gains in speed of verbal processing by infants in the 2nd year. *Psychological Science 9,* 228-231.

Fernald, A., & McRoberts, G. (1991). Prosody and early lexical comprehension. Paper presented at the meeting of the Society for Research on Child Development. Seattle.

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

Fernald, A. (1985). Four-month old infants prefer to listen to motherese. *Infant Behavior and Development, 8, 181-195.*

Fischer, S.D., Delhorne, L.A., & Reed, Ch.A. (1999). Effects of rate of presentation on the reception of American Sign Language. *Journal of Speech, Language, and Hearing Research 42*, 568-582.

Foulke, E. (1971). The perception of time-compressed speech. In: D.L. Horton & J.J. Jenkins (Eds.), *The perception of language* (pp. 79-101). Columbus, OH: Merrill.

Foulke, E., & Sticht, T. G. (1969). Review of research on the intelligibility and comprehension of accelerated speech. *Psychological Bulletin 72,* 50-62.

Goldfield, B.A., & Reznick, J.S. (1990). Early lexical acquisition: rate, content, and the vocabulary spurt. *Journal of Child Language 17,* 171-183.

Goldfield, B.A., & Reznick, J.S. (1996). Measuring the vocabulary spurt: a reply to Merivs & Bertrand. *Journal of Child Language 23,* 241-246.

Gordon-Salant, S., & Fitzgibbons, P.J. (1999). Profile of auditory temporal processing in older listeners. *Journal of Speech, Language and Hearing Research*, 42 (2), 300-311.

Gordon-Salant, S., & Fitzgibbons, P.J. (1993). Temporal factors and speech recognition performance in young and elderly listeners. *Journal of Speech and Hearing Research, 36,* 1276-1285.

Grosjean, F. (1980). Spoken word recognition processes and the gating paradigm. *Perception and Psychophysics, 38,* 299-310.

Grosjean, F. (1985). The recognition of words after their acoustic offset: evidence and implications. *Perception and Psychophysics, 38,* 299-310.

Haith, M.M., Wentworth, N. & Canfield, R.L. (1993). The formation of expectations in early infancy. *Advances in Infancy Research* 8, 251-297.

Hood, B.M., & Atkinson, J. (1993). Disengaging visual attention in the infant and adult. *Infant Behaviour and Development* 16, 405-422.

Jusczyk, P.W. , Friederici, A.D., Wessels, J.M.I., Svenkerud, V., & Jusczyk, A.M. (1993). Infant's sensitivity to the sound pattern of native-language words. *Journal of Memory and Language* 32, 402-420.

Jusczyk, P. W. (1986). Towards a model for the development of speech perception. In J. Perkell, & D.H. Klatt (Eds.), *Invariance and variability in speech processes* (pp.1-19). Hillsdale, NJ: Erlbaum.

Kahneman, D. (1973). Attention and effort. Englewood Cliffs, NJ: Prentice-Hall.

King, P.E., & Behnke, R.R. (1989). The effect of time-compressed speech on comprehension, interpretive, and short-term listening. *Human Communication Research* 15, 428-443.

Lane, H., & Grosjean, F. (1973). Perception of reading rate by speakers and listeners. *Journal of Experimental Psychology 97*, 141-147.

Leonard, C.L., Baum, S.R., & Pell, M.D. (2000). Context use by right-hemisphere-damaged individuals under a compressed speech condition. *Brain and Cognition, 43 (1-3),* 315-319.

Marchman, V., & Bates, E. (1994). Continuity in lexical and morphological development: a test of the critical mass hypothesis. *Journal of Child Language 21,* 339-366.

Marslen-Wilson, W.D. (1984). Function and process in spoken word recognition: A tutorial review. In: H. Bouma, & D. Bouwhuis (Eds.), *Attention and performance X: Control of language processes (*pp. 125-148). Hillsdale, NJ: Erlbaum.

Marslen-Wilson, W.D. (1987). Functional parallelism in spoken word-recognition. *Cognition, 25*, 71-102.

Marslen-Wilson, W.D. (1993). Issues of process and representation in lexical access. In G. T. Altmann, & R.Shillcock (Eds.), *Cognitive models of speech processing: The second Sperlonga meeting* (pp. 187-211). Hillsdale, NJ: Erlbaum.

Mehler, J., Sebastian, N., Altmann, G., Dupoux, E., Christophe, A., & Pallier, Christophe (1993). Understanding compressed sentences: The role of rhythm and meaning. In P. Tallal & A.M. Galaburda (Eds.), *Temporal information processing in the nervous system: Special reference to dyslexia and dysphasia* (pp. 272-282).

Mehler, J., & Christophe, A. (1995). Maturation and Learning of Language in the first year of life. In M. Gazzaniga, (Ed.), *The cognitive neurosciences.* (pp. 943-958), Cambridge, London: MIT Press.

Mervis, C.B., & Bertrand, J. (1995). Early lexical acquisition and the vocabulary spurt: a response to Goldfield & Reznick. *Journal of Child Language 22,* 461-468.

Mills, D.L., Werker, J.F., Neville, H., Pratt, C., & Zangl, R. Semantic processing of phonetically similar words. (in preparation).

Mills, D.L., Thal, D., Llamas, T., & Pratt, C. Brain organization and expressive language delay: Electrophysiological evidence from late talkers and children with a family history of language impairment (submitted). In B. Wulfeck & J. Reilly, (Eds.), Plasticity and development: Language in atypical children. Special issue, *Brain and Language* (under review).

Mills, D.L. (1999, April). Cerebral specialization before and after the vocabulary spurt: An event-related potential study of novel word learning. Paper presented in symposium: *Early vocabulary growth: mechanisms of change.* Biennal Meeting for Research in Child Development. Albuquerque, NM.

Mills, D.L., Coffey-Corina, S.A., & Neville, H.J. (1997). Language Comprehension and Cerebral Specialization from 13 to 20 Months. *Developmental Neuropsychology, 13(3), 397-445.*

Mills, D.L., Coffey-Corina, S.A. & Neville, H.J. (1993). Language acquisition and cerebral specialization in 20-month-old Infants. *Journal of Cognitive Neuroscience, 5/3, 317-334.*

Miyake, A., Carpenter, P., & Just, M. (1994). A capacity appraoch to syntactic comprehension disorders: Making normal adults perform like aphasic patients. *Cognitive Neuropsychology, 11*, 671-717.

Munson, B. (in press). Relationships between vocabulary size and spoken word recognition in children aged 3-7. *Contemporary Issues in Communication Science and Disorders.*

Nooteboom, S.G., & Doodeman, G.J.N. (1984). Speech quality and the gating paradigm. In M.P.R. Van den Broek & A. Cohen, (Eds.), *Proceedings of the Tenth International Congress of Phonetic Sciences* (pp. 481-485). Dordrecht: Foris.

Norman, D.A., & Bobrow, D.G. (1975). On data-limited and resource-limited processes. *Cognitive Psychology* 7, 44-64.

Obler, L.K., Fein, D., Nicholas, M., & Albert, M.L. (1991). Auditory comprehension and aging: decline in syntactic processing. *Applied Psycholinguistics* 12, 433-452.

Pashler, H.E. (1998). *The psychology of attention.* Cambridge, MA: MIT Press.

Poldrack, R.A., Protopapas, A., Nagarajan, S., Tallal, P., Merzenich, M., Temple, E., & Gabrieli, J.D. (1998). Auditory processing of temporally compressed speech: An fMRI study. Presented at the Cognitive and Neuroscience Society, Fifth Annual Meeting, San Francisco, CA.

Reznick, J.S., & Goldfield, B.A. (1992). Rapid change in lexical development in comprehension and production. *Developmental Psychology 28(3),* 406-413.

Roberts, J.A. (1998). Cognitive and linguistic development before and after the vocabulary spurt. Dissertation. Dissertation Abstracts International: Section B: The Sciences & Engineering, 59 (3-B), p. 1393.

Schafer, G. (1998). Word learning in infancy. Doctoral dissertation, Oxford University.

Schafer, G., & Plunkett, K. (1998). Rapid word learning by fifteen-month-olds under tightly controlled conditions. *Child Development 69,* 309-320.

Schmitt, J.F., & Carroll, M.R. (1985). Older listeners' ability to comprehend speaker-generated rate alteration of passages. *Journal of Speech and Hearing Research* 28, 309-312.

Shannon, R.V., Zeng, F-G., & Wygonski, J. (1998). Speech recognition with altered spectral distribution of envelope cues. *Journal of the Acoustical Society of America, 104/4,* 2467-2476.

Speer, S.R., Wayland, S.C., Kjelgaard, M.M., & Wingfield, A. (1994). Effect of speaking rate and paragraph structure on the producation of sentence prosody. *Journal of the Acoustical Society of America 95,* 2979.

Stark, R. & Montgomery, James (1995). Sentence processing in language-impaired children under conditions of filtering and time compression. *Applied Psycholinguistics* 16 (2), 137-154.

Sticht, T.G., & Gray, B.B. (1969). The intelligibility of time compressed words as a function of hearing loss. *Journal of Speech and Hearing Research 12,* 443-448.

Stine, E.A.L., Wingfield, A., & Myers, S.D. (1990). Age differences in processing information from television news. The effects of bisensory augmentation. *Journal of Gerontology: Psychological Sciences 45,* P1-P8.

Stine, E.A.L., Wingfield, A. & Poon, L.W. (1986). How much and how fast: rapid processing of spoken language in later adulthood. *Psychology and Aging* 1, 303-311.

Stuart, A. & Phillips, D. (1996). Word recognition in continuous and interrupted broadband noise by young normal-hearing, older normal-hearing, and presbyacusic listeners. *Ear and Hearing* 17, 478-489.

Swingley, D., & Fernald, A. (2000). Visual context and word recognition at 24 months (Manuscript under review).

Swingley, D., & Aslin, D. (2000). Spoken word recognition and lexical representation in very young children. *Cognition* 76, 147-166.

Swingley, D., Pinto, J.P., & Fernald, A. (1999). Continuous processing in word recognition at 24 months. *Cognition* 71, 73-108.

Swingley, D., Pinto, J.P., & Fernald, A. (1998). Assessing the speed and accuracy of word recognition in infants. In: Rovee-Collier, C., Lipsitt, L.P. & Hayne, H. (Eds.), *Advances in infancy research* (Vol. 12, pp. 257-277). Stamford, CT: Ablex.

Tun, P., Wingfield, A., Stine, E.A.L., & Mecsas, C. (1992). Rapid speech processing and divided attention: Processing rate versus processing resources as an explanation of age effects. *Psychology and Aging, 7,* 546-550.

Tyler, L. (1984). The structure of the initial cohort: Evidence from gating. *Perception and Psychophysics, 36,* 417-427.

Utman, J. A., & Bates, E. (2000). Effects of acoustic distortion and semantic context on lexical access. (under review).

Vaughan, N.E., & Letowski, T. (1997). Effects of age, speech rate, and type of test on temporal auditory processing. *Journal of Speech, Language and Hearing Research 40/5,* 1192-1200.

Walley, A.C. (1993). The role of vocabulary development in children's spoken word recognition and segmentation ability. *Developmental Review* 13, 286-350.

Washburn, D.A., & Putney, R.Th. (2001). Attention and Task Difficulty: When is performance facilitated? *Learning and Motivation* 32, 36-47.

Weismer, S.E., & Hesketh, L.J. (1996). Lexical learning by children with specific language impairment: effects of linguistic input presented at varying speaking rates. *Journal of Speech and Hearing Research 39/1,* 177-190.

Wingfield, A., Tun, P.A., Koh, Ch.K., & Rosen, M.J. (1999). Regaining lost time: Adult aging and the effect of time restoration on recall of time-compressed speech. *Psychology and Aging 14,* 380-389.

Wingfield, A., & Ducharme, J.E. (1999). Effects of age and passage difficulty on listening-rate preferences for time-altered speech. *Journals of Gerontology: Series B: Psychological Sciences & Social Sciences 54/3,* P199-P202.

Wingfield, A. (1996). Cognitive factors in auditory performance: Context, speed of processing, and constraints of memory. *Journal of the American Academy of Audiology 7,* 175-182.

Table 1: List of 24 target words used in experiments 1 and 2

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 2: ADULT REACTION TIMES in ms for each of the five perceptual conditions (level of distortion, acoustic shape of the auditory word). Note: * Unaltered words were attributed to either severe/moderate level for reasons of counterbalancing; they did not differ in their acoustic shape.

Auditory Condition	Mean RT in ms
Severe, time-compressed	310.12
Severe, low-pass filtered	38.34
Severe, unaltered*	–335.15
Moderate, time-compressed	43.4
Moderate, low-pass filtered	–143.78
Moderate-unaltered*	–364.07

Table 3: Total number of 95 infants broken down for each age level

Age in months	Number of subjects	Age in months	Number of subjects
12	N=13	21	N=6
13	N=6	22	N=6
14	N=3	23	N=4
15	N=7	24	N=11
16	N=8	25	N=2
17	N=3	26	N=1
18	N=7	27	N=1
19	N=11	30	N=2
20	N=3	31	N=1

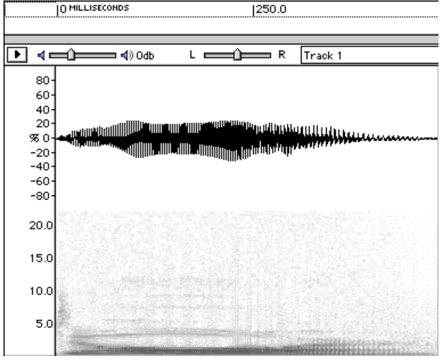
Table 4: EPOCH ACCURACY – D+T-TRIALS. A summary of the ANOVA values (degrees of freedom, F-values and p-values) for the two- and three-way interactions for target fixations including both distractor and target trials. The statistics are grouped for a. age and b. vocabulary. (Note: * = significant)

a. Age	**df**	**F-value**	**p-Value**	**b. Vocabulary**	**df**	**F-value**	**p-value**
Condition × Age	6	2.21	.04*	Condition × Vocabulary	6	1.44	.20
Condition × Epoch	4	7.04	.0001*	Condition × Epoch	4	7.32	.0001*
Epoch × Age	6	5.34	.0001*	Epoch × Vocabulary	6	5.34	.0001*
Condition × Epoch × Age	12	1.95	.03*	Condition × Epoch × Vocabulary	12	2.51	.004*

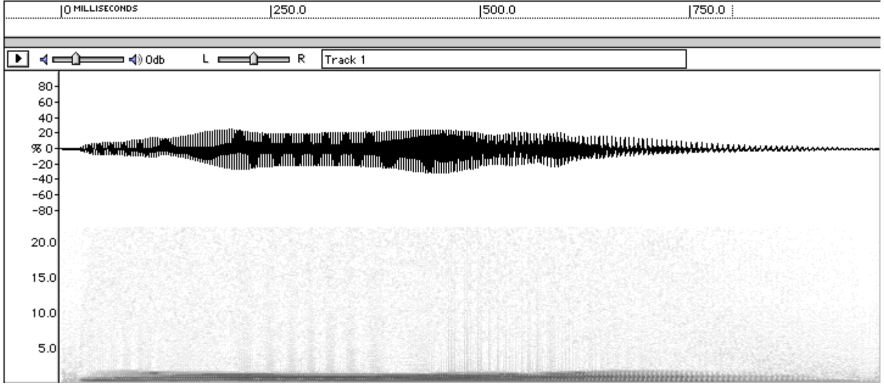
Table 5: EPOCH ACCURACY – T-TRIALS. A summary of the ANOVA values (degrees of freedom, F-values and p-values) for the two- and three-way interactions for target fixations measured for target trials only. The statistics are grouped for a. age and b. vocabulary. (Note: * = significant)

a. Age	**df**	**F-value**	**p-Value**	**b. Vocabulary**	**df**	**F-value**	**p-value**
Condition × Epoch	4	5.64	.0002*	Condition × Epoch	4	6.58	.0001*
Condition × Epoch × Age	12	2.81	.0011*	Condition × Epoch ×Vocabulary	12	2.52	.0034*

door: 50% time-compressed



door: 1.5kHz low-pass filtered



door: unaltered

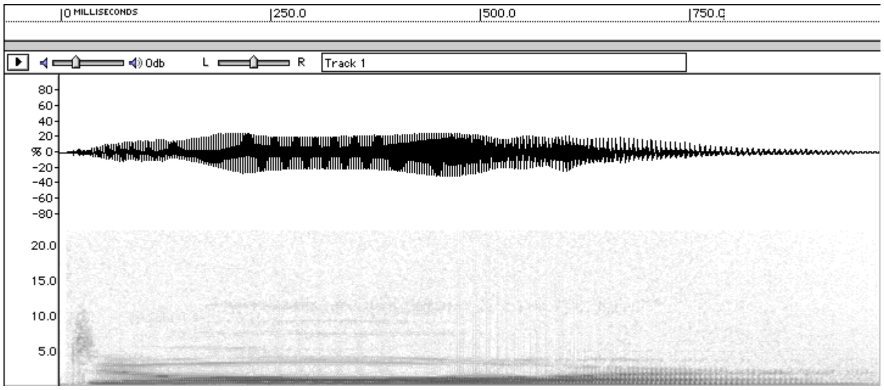


Figure 1: Sample of wave and spectrum for the target word ‘*door*’ in time-compressed (top panel), low-pass filtered (center panel) and unaltered (bottom panel) condition

Fig. 2A: Epoch Accuracy, D+T-trials, Age

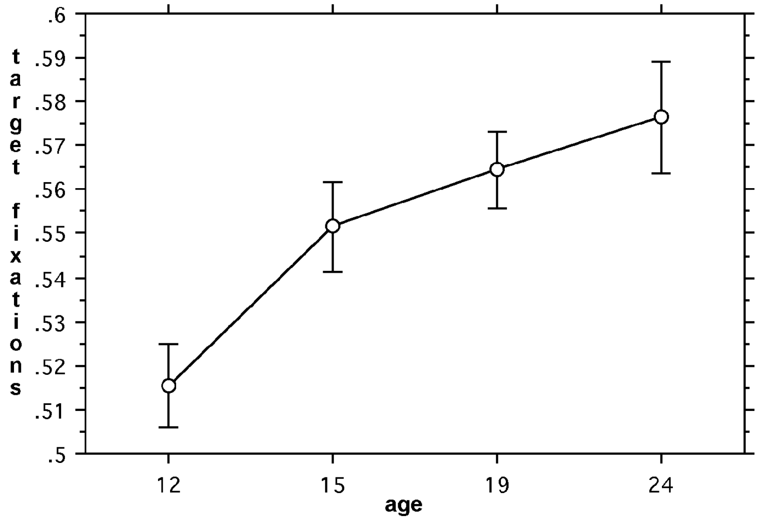
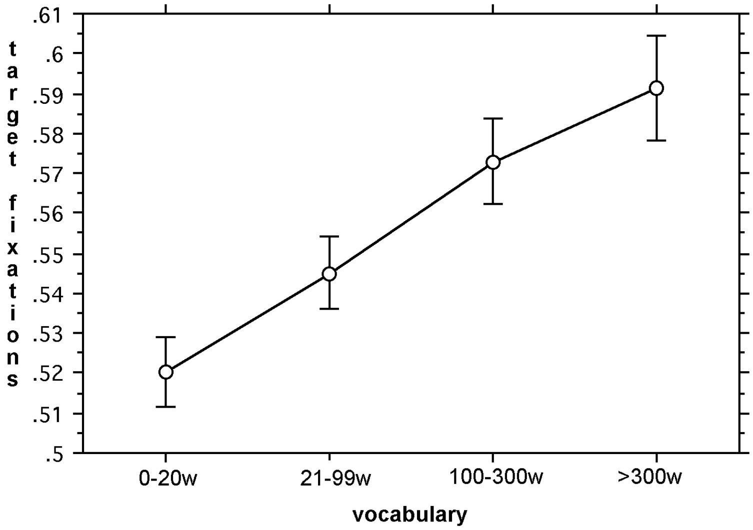


Fig. 2B: Epoch Accuracy, D+T-trials, Vocabulary



Effects of Age and Vocabulary. The developmental curves reflect the proportion of correct target fixations once grouped by age (2A), the other time by vocabulary (2B), including both distractor and target trials.

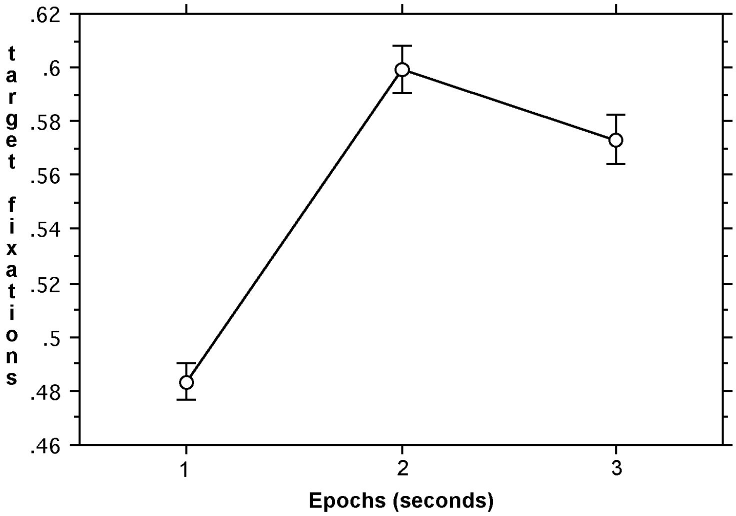


Fig. 3: Epoch Accuracy, D+T-trials, target fixation over epochs

The proportion of correct target fixations are plotted separately for each epoch (1st, 2nd and 3rd) for distractor and target trials combined.

Fig. 4A: Perceptual Condition x Age

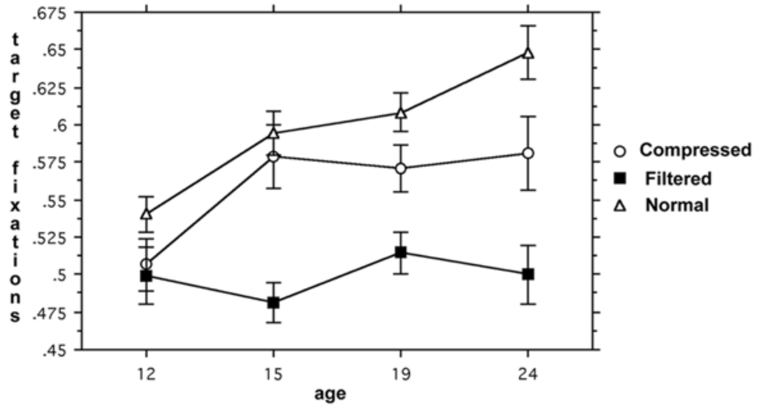


Fig. 4B: Perceptual Condition x Epoch

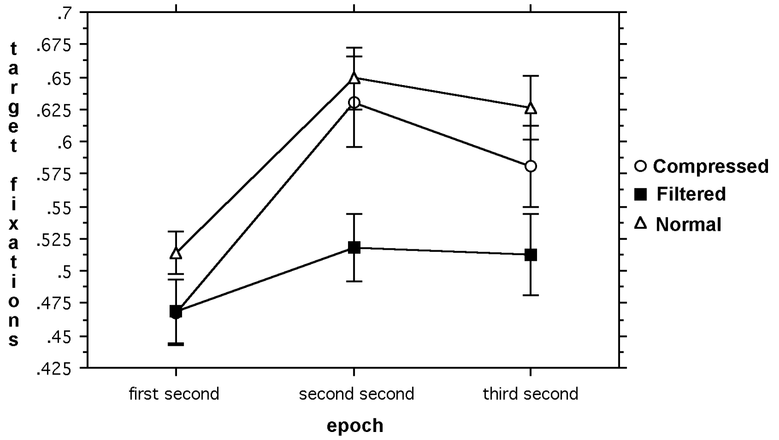
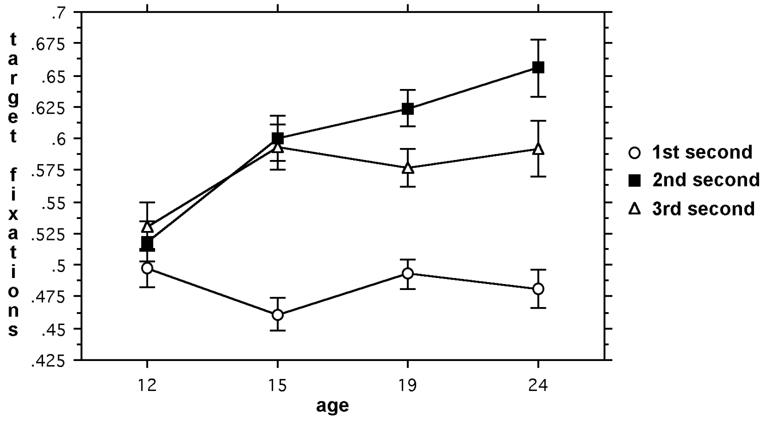
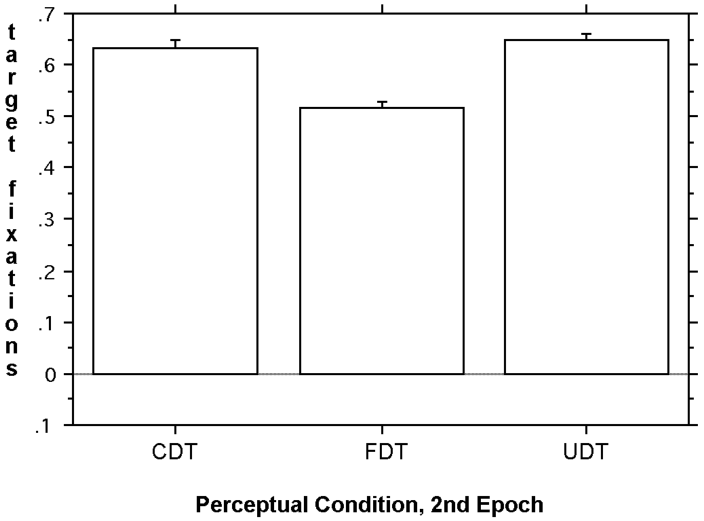


Fig. 4C: Epoch x Age



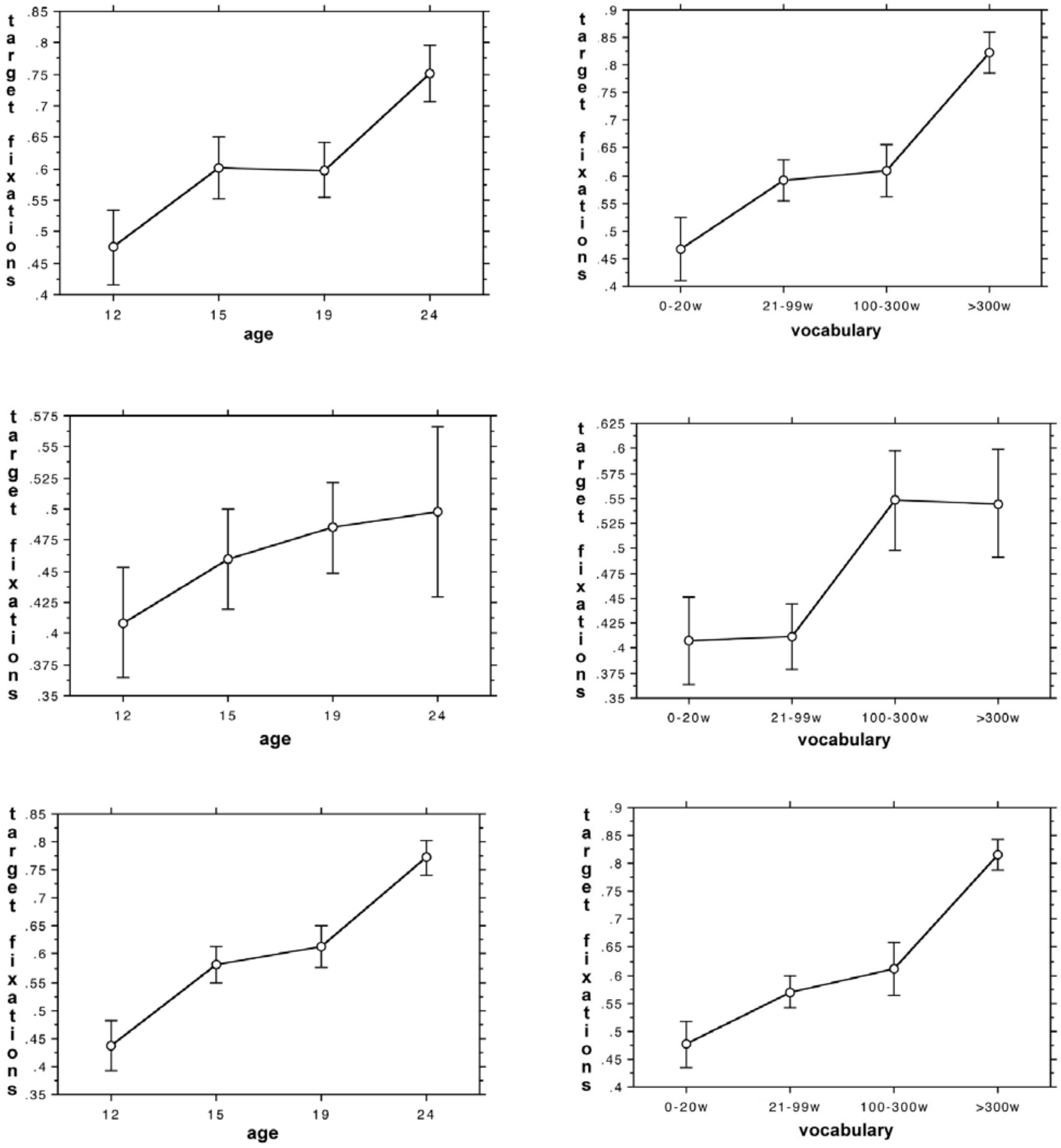
EPOCH ACCURACY – D+T-TRIALS, 2-WAY-INTERACTIONS for Perceptual Condition × Age (4A), Perceptual Condition × Epoch (4B) and Epoch × Age (4C) for distractor and target trials combined.

Fig. 5: Epoch Accuracy, D+T-trials, 2nd Epoch



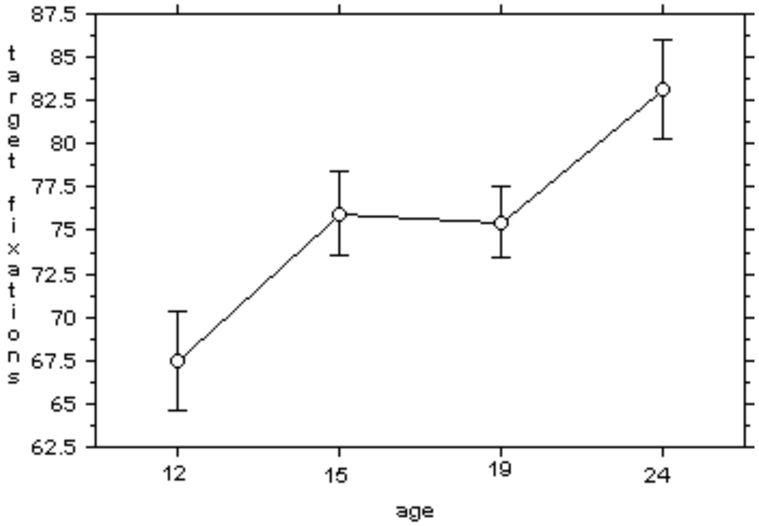
Effect of Perceptual Condition. The proportion of correct target fixations is presented in the 2nd Epoch (only) for each of the conditions – unaltered, time-compressed and low-pass filtered – including both distractor and target trials

Figures 6A-6C



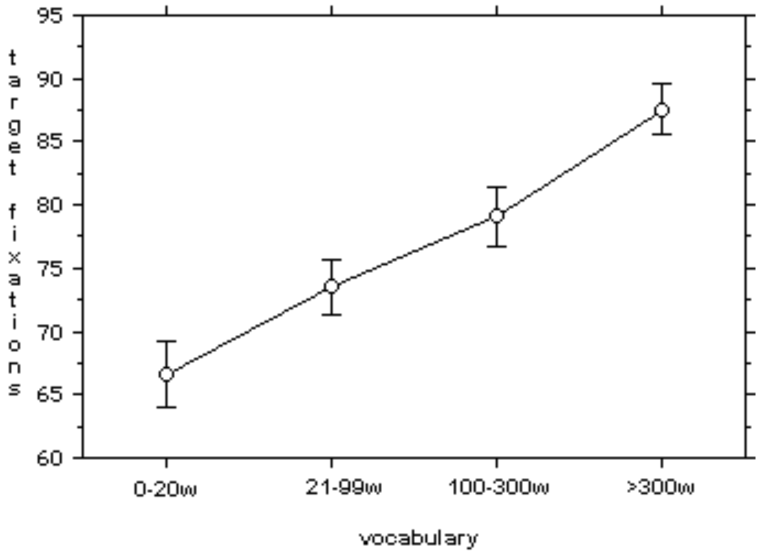
Effect of Age/Vocabulary for each condition.The proportion of correct target fixations is presented once for an age and once for a vocabulary grouping for each of the three conditions - time-compressed (top panel), low-pass filtered (center panel) and unaltered (bottom panel). Target fixations reflect measurements for the 2nd Epoch only. Age groupings are presented on the left, vocabulary groupings on the right (from top to bottom respectively).

Fig. 7: Highest Peak Amplitude, Age group



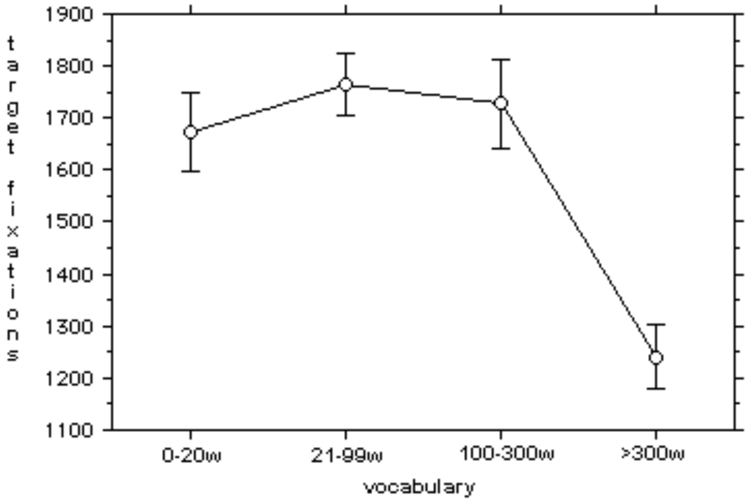
Effect of Age. The developmental curve presents the peak amplitudes (= % target fixations) for each of the four age groups (12, 15, 19 and 24 months).

Fig. 8: Highest Peak Amplitude, Vocabulary group



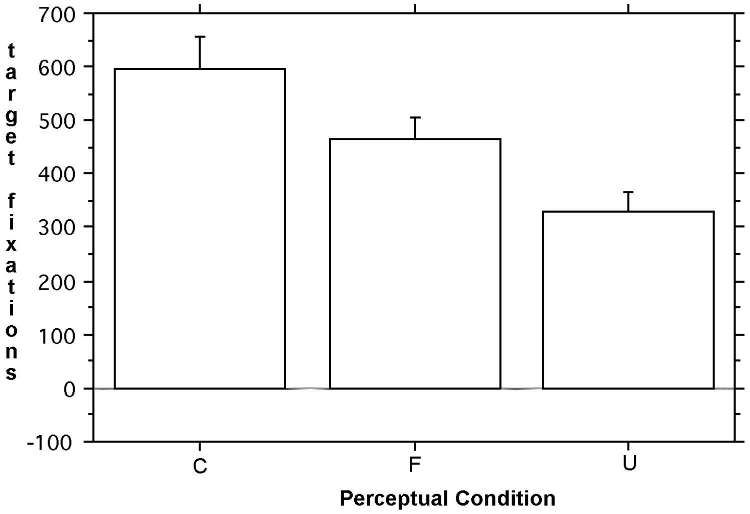
Effect of Vocabulary. The developmental curve presents the peak amplitudes (=%target fixations) for each of the four vocabulary groups (0-20, 21-99, 100-300 and > 300 words).

Fig. 9: Peak Latency, Vocabulary Group



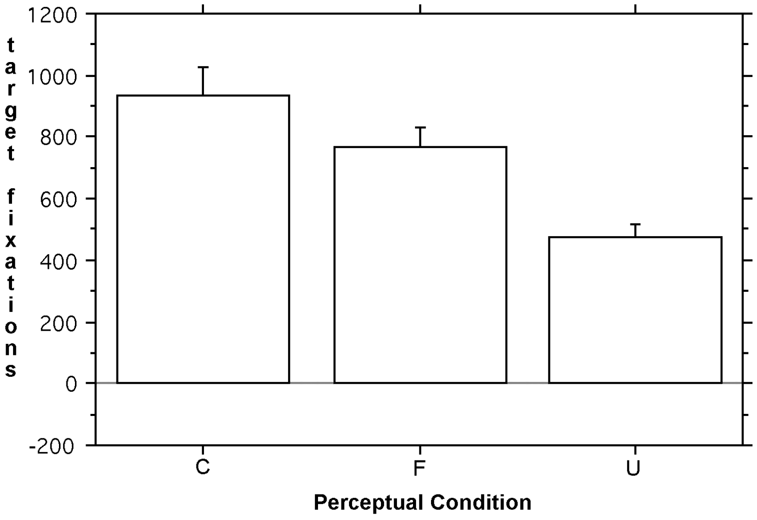
Effect of Vocabulary. The developmental curve presents the peak latencies in msec (= reaction times in ms to the highest peak amplitude) for each of the four vocabulary groups (0-20, 21-99, 100-300 and > 300 words).

Fig. 10: Peak Duration, D-trials, Target fixation



Effect of Perceptual Condition. The bar graphs present the target fixation times for unaltered, time-compressed and low-pass filtered trials calculated from the onset to the offset of the highest peak.

Fig. 11: Peak Duration, T-trials, Target fixation



Effect of Perceptual Condition. The bar graphs present the total of target fixation times for unaltered, time-compressed and low-pass-filtered trials calculated from the onset to the offset of 100% target-looking time.