Building a word-form lexicon in the face of variable input: Influences of pitch and amplitude on early spoken word recognition

Leher Singh

Boston University

Katherine S. White & James L. Morgan

Brown University

Address for correspondence:

Leher Singh
Dept. of Speech, Language & Hearing Sciences
Boston University
635 Commonwealth Avenue
Boston, MA 02215
(617) 353 7491
leher@bu.edu

Katherine White & James L. Morgan
Box 1978, Dept. of Cognitive and Linguistic Sciences
Brown University
Providence, RI 02912

Acknowledgements: This project was supported by a grant from the National Institutes of Health (5 RO1 HD32005) to J. M. and a grant from the National Institutes of Health to LS (5 R03 HD046676) to L.S. We thank Amanda Casile, Sarah Nestor, Eileen Leung, Karen Rathbun, and Lori Rolfe for assistance with data collection.

Abstract

To build their first lexicon, infants must first be able to recognize words in the input. This task is made challenging by the inherent variability of speech. Potential sources of variability include changes in speaker identity, vocal emotion, amplitude and pitch. English-speaking adults can recognize a word regardless of these changes and mature word recognition is not impeded by changes in amplitude or pitch. In this set of studies, we independently manipulate amplitude and pitch to examine whether infants' lexical processing is similarly invulnerable to changes in surface form. We found that 7.5 month-old infants at the earliest stages of word recognition can recognize a word if it is presented in a different amplitude, but not in a different pitch. By 9 months, infants are able to recognize words independent of changes in pitch and amplitude, thus appearing to appreciate the irrelevance of both properties in determining lexical identity. Results are interpreted with respect to why infants may treat pitch and amplitude distinctly in spoken word recognition.

Building a word-form lexicon in the face of variable input:
Influences of pitch and amplitude on early spoken word recognition

Building a lexicon is one of the most fundamental tasks in learning a language. While we continue to learn words even as adults, the construction of an initial lexicon in infancy is uniquely challenging. As adults, we know many facts about words, including what they mean, how their forms can be systematically modified to impart changes in meaning, how they can be combined into larger units to form phrases, and how these linguistic forms are conventionally used in the language community. Infants learning words do not have *a priori* access to this information. As a result, they must induce the meanings and grammatical properties of words from their input.

Before they can attach meaning to words, however, infants must be able to categorize different exemplars (i.e. tokens) as instances of the same word. This is only possible if they can recognize the lexical equivalence of tokens in different contexts. Recognition across contexts is a non-trivial task, as the linguistic and extra-linguistic context in which a word appears alters its phonetic realization, making each new token physically distinct from previous tokens. Phonetic variation caused by context is rampant: A word can be produced by talkers of different ages or of either gender. Even within a talker there are many differences (e.g. in vocal affect, speech rate, emphasis, etc.) across productions. As adults, we know which of these changes are lexically relevant and which are not. However, in order to develop stable lexical representations, infants must learn to separate sources of phonetic variability that do not affect lexical identity from sources that do have an impact on lexical identity. In this article, we explore infants' capacities for recognizing the lexical equivalence of words when non-meaningful phonetic variation is introduced, and we consider whether two types of variation (amplitude and pitch) are treated equally at the earliest stages of lexical development.

By 6 months, infants can segment and recognize words in fluent speech when they occur adjacent to familiar names (Bortfeld, Morgan, Golinkoff, & Rathbun, 2005), and by 7.5 months, this ability has generalized to novel words (Jusczyk & Aslin, 1995). However, acoustic dissimilarity across tokens, whether lexically relevant or not, disrupts infants' recognition of familiarized words, presumably because they have yet not sorted dimensions of sound that communicate meaning from those that do not. For example, when there is a change between familiarization and test in the gender of the speaker or in a single speaker's affect or emphatic stress, infants at 7.5 months fail to recognize repetitions of familiarized words (Bortfeld & Morgan, in submission; Houston, 2000; Houston & Jusczyk, 2000; Singh, Morgan & White, 2004).

In a series of studies probing the effects of talker variability on word recognition, Houston (2000; Houston & Jusczyk, 2000) familiarized infants with words produced by one talker. During the recognition phase of the experiment, infants heard these words spoken either by the same talker or by another talker (with a distinct voice or of the opposite gender). At 7.5 months, infants recognized the words only when the talker remained constant across the familiarization and recognition phases. Thus, infants at this age were unable to disregard acoustic information indexing talker and required high perceptual similarity across tokens for recognition. Singh, Morgan & White (2004) extended this finding to productions from a single talker that varied in talker affect. Seven and a half month-olds did not recognize words spoken in happy affect when they occurred later in neutral passages (and vice versa). Bortfeld & Morgan (in submission) showed further that changes from emphatic to non-emphatic stress (or vice versa) between familiarization and recognition testing similarly impedes 7-month-olds' word recognition. These demonstrations of *perceptual matching* indicate that paralinguistic or non-linguistic variations seriously disrupt young infants' word recognition. Later in development, by 10.5 months however, infants are able to recognize lexical equivalence (according to the criterion used in the Headturn Preference Procedure) despite the changes in surface form caused by either speaker identity or speaker affect (Houston & Jusczyk, 2000; Singh et al., 2004).

The acoustic cues specifying talker identity, or accompanying the expression of vocal affect, emphasis, or speech rate are many. Among the acoustic properties that have been highlighted as important for talker identification are formant range, variation in fundamental frequency (F0), and the shape of the laryngeal spectrum, all of which contribute to the perception of voice quality (see Remez, Fellowes & Rubin, 1997). Vocal affect is encoded by variation along a number of dimensions (Murray & Arnott, 1993; Scherer, 1986; Williams & Stevens, 1972); in particular, it can be reliably identified using fundamental frequency (F0), contours (Scherer, 1986; Williams & Stevens, 1972), and the proportion of high frequency energy in the spectrum (Banse & Scherer, 1996). Fundamental frequency (F0) is also an important correlate of emphatic stress, along with amplitude and duration (Bortfeld & Morgan, in submission). It is important to note that some of these acoustic cues- F0, in particular-are employed to signal lexical distinctions in many languages of the world. In tonal languages, for example, variations in F0 contours are phonemic; in stress-based languages, such as

English, changes in F0, amplitude, and duration can mark lexical stress (e.g. perFECT vs PERfect). Therefore, variation that is meaningful and is therefore part of the signal in some contexts might constitute noise in other contexts. Similarly, the relationship between acoustic variation and meaning varies considerably across languages, making it incumbent on an early learner to distinguish signal and noise in a language- and context-sensitive fashion.

The presence of surface variation clearly burdens spoken word recognition in infants. However, it also imposes a cost on adults in language processing tasks. Evidence has accumulated that word recognition in adults is affected by the acoustic/phonetic realizations of words. Adults show sensitivity to changes in vocal affect and talker gender in both explicit memory tasks and implicit tasks of spoken word recognition (e.g. Bradlow, Nygaard & Pisoni, 1999; Church & Schacter, 1994; Luce & Lyons, 1998; Nygaard, Sommers & Pisoni, 1995; Palmeri, Goldinger & Pisoni, 1993; Sommers, Nygaard & Pisoni, 1994). All types of surface variation do not affect spoken word recognition in adults equally, however. In particular, although changes in talker identity, speech rate, and fundamental frequency all have demonstrable effects on adult spoken word recognition, changes in amplitude do not (Bradlow et al., 1999; Church & Schacter, 1994; Sommers et al., 1994).

This raises the question of whether infants are also exceptional in their treatment of amplitude or whether processing in infancy is disrupted by surface properties that reliably characterize their encounters with a particular word. It stands to reason that infants, lacking knowledge of native acoustic-phonemic correspondences, may consider any property that emerges across repetitions of a word to be relevant to its identity, whether that property is speech rate, vocal affect, amplitude or any other commonly varying factor. For example, if a particular word is always spoken with happy affect, infants may assume that the acoustic correlates of happy affect are as relevant to the identity of the word as its phonological form. On this view, any surface property that covaries with the phonological properties of a word is likely to be judged as a defining characteristic of that word for a young learner. This leads to successful word segmentation only in cases where there is an *acoustic match* across exemplars, as observed in infants at 7.5 months. It may be the later process of learning word meanings that highlights precisely which changes in surface form are lexically relevant. This type of 'covariance' account is certainly consistent with infants' treatment of words amidst changes in talker gender and vocal affect.

A potential test of this hypothesis would be to determine how factors that do not impact upon adult lexical processing (i.e. amplitude) are treated by infants. If infants are apt to treat any covarying property of a word as relevant to its identity, even changes in amplitude (which adults appear to normalize) would disrupt processing in infancy. Experiment 1 is designed to test this prediction by investigating the consequences of amplitude changes, to which adults appear invulnerable, on spoken word recognition at 7.5 months.

EXPERIMENT 1

In this study, infants were tested on their capacity to recognize words in fluent speech when the tokens differed in amplitude. Infants were familiarized with two words, one at relatively high amplitude and another at relatively low amplitude. Infants then heard passages with sentences containing the familiarized words, as well as comparable passages containing nonfamiliarized words. Amplitude differences between the stimulus sets were approximately 21 dB, about the difference between quiet conversation and an outboard motor. Therefore, the lower amplitude stimuli were soft in volume, although clearly audible. The soft volume was not considered a problem during familiarization, when infants were presented with multiple tokens of words in isolation and simply had to commit the words to memory. However, processing demands are higher during the recognition phase, where infants have to track, segment and classify the target words. Moreover, the natural amplitude variation across an entire sentence may have rendered parts of the low amplitude sentences inaudible. Therefore, amplitude was manipulated only during familiarization. All recognition passages were presented at the higher amplitude.

<u>Participants</u> Sixteen full-term, English-exposed 7.5-month-olds participated in the study (12 males and 4 females), recruited from Massachusetts birth records and advertisements. While this is not an ideal balance of males and females, there have never been any reported effects of gender in other studies on infant word segmentation and therefore, we do not expect such effects in the present study. Mean age of participants was 232 days (range = 215 days to 250 days). Data from one additional infant were not included for inattention.

<u>Stimuli</u> Stimuli for Experiment 1 consisted of four monosyllabic words (*bike*, *hat*, *tree* and *pear*) and four six-sentence passages recorded by a mother addressing her infant. When recording the stimuli, the speaker was asked to use infant-directed speech for all tokens and to communicate positive affect. The stimuli used were the ’happy’ stimuli

used in Singh et al. (2004), except that they were transformed to create high and low amplitude tokens.

For familiarization stimuli, thirty tokens of each word were recorded (for recording details and stimulus selection criteria, see Singh et al., 2004). For each word type, 15 tokens were selected to serve as familiarization stimuli based on the clarity of the stimuli. Recognition stimuli consisted of sentences containing the target words (*bike*, *hat*, *tree*, and *pear*). All passages are shown in the Appendix. Recognition stimuli consisted of 24 sentences containing the target words (*bike*, *hat*, *tree*, and *pear*). Each target word appeared in a 6-sentence passage, once per sentence. Within each passage, the target word appeared twice each in initial, medial and final sentence positions (see Appendix).

Amplitude variation was introduced by using the original set of recordings as High Amplitude stimuli. A matched Low Amplitude set was created using the BLISS program (Mertus, 2002), by decreasing the amplitude of the stimuli by approximately 21dB. This difference in amplitude was chosen to fit with previous studies investigating effects of amplitude variation on adult word recognition. In these studies, it was demonstrated that differences ranging from 10 to 30 dB had no effect on word identification in adults (Nygaard, Sommers & Pisoni, 1995; Sommers, Nygaard & Pisoni, 1994). A relatively large change in amplitude was selected here in order to ensure that the stimuli were sufficiently perceptually dissimilar for the infants.

Acoustic measures were obtained for both stimulus sets using Praat (Boersma & Weenink, 1996). Both pitch and duration measures were identical across the two stimulus sets. All acoustic measurements are included in Table 1. Similar to previous studies investigating infant word recognition (e.g., Houston & Jusczyk, 2000; Jusczyk & Aslin, 1995; Singh, et al., 2004), infants heard citation form tokens of two words during familiarization. Half of the infants heard the words *bike* and *hat* while the other half heard *tree* and *pear*. For each infant, one word was heard at High Amplitude and the other at Low Amplitude. The words were used as familiarization stimuli and the assignment of amplitude to words was counterbalanced across subjects. During recognition testing, infants heard passages containing all four words. As a result of counterbalancing, familiar passages for some infants were unfamiliar to others and vice versa. All infants heard passages spoken at High Amplitude.

Apparatus Testing was conducted in a three-walled testing booth within a sound-proofed anechoic chamber. Each wall of the booth was 120 cm wide. A chair was positioned at the open end of the booth where the parent sat with the infant on his/her lap. The infant sat approximately 110 cm from the front of the booth. Loudspeakers were located behind both side walls of the booth. At the infants' eye level, 86 cm above the floor, a white light was mounted on the front wall. Each of the side walls had a similar blue light at the same level. A CCTV camera was mounted behind the testing booth 12.3 cm above the white light. In a separate control room, a television monitor was connected to the CCTV camera in the testing booth. Participants were displayed on the monitor in the control room, where the experimenter judged infants' looking, pressing buttons on the mouse of a Windows computer to control the customized experimental software. The computer was equipped with a Sound-Blaster compatible soundboard connected to the amplified speakers.

Procedure Infants were tested using the Headturn Preference Procedure (HPP) (Kemler Nelson, Jusczyk, Mandel, Myers, Turk, & Gerken, 1995). The infant was seated on the parent’s lap facing the center light. The parent listened to instrumental music over headphones to mask the stimuli. Each trial began with the center light flashing until the experimenter judged that the infant fixated on the flashing light. At that point, this light was turned off, and one of the side lights began to flash to attract the infant’s attention to the side. Side of presentation was randomized across trials, so that all stimuli occurred on both sides. After the infant turned to look at the flashing side light, the speech stimuli for that trial began to play. The sound continued to play and the side light remained on for the duration of the infant’s fixation on the light. Each trial continued until the infant looked away for two seconds, or until 20 seconds of looking time had been accumulated during that trial. If the infant looked away, but then looked back within two seconds, the trial continued. If the infant’s total looking time was below 2 seconds, the trial was repeated with a new randomization of the trial stimuli; otherwise, the procedure advanced to the next trial.

Familiarization began with trials alternating between the two target words. Once the infant had exceeded 30 seconds of looking time with one word, all subsequent familiarization trials presented the alternate word. This modification of the HPP was instituted to ensure that differences in looking times during recognition testing could not be due to different amounts of familiarization with the two target words. When the infant reached 30 seconds of looking time with the second word, the test phase began. The words used as familiarization stimuli and the assignment of amplitude to words were counterbalanced across subjects. As a result of this design, across subjects each item served every possible role (matched familiarization word/mismatched familiarization word/unfamiliar word).

Recognition testing consisted of four blocks of trials, each block containing one passage. Each passage comprised six sentences that contained one of the four words. Therefore, within each block, infants had an opportunity to listen to passages containing ‘bike’, ‘hat’, ‘tree’, and ‘pear.’ Within each block of four passages, the order of passages was randomized for each infant. In addition, the order of sentences within passages was also randomized on each trial.

The test procedure was similar to the familiarization procedure, except that the side light continued to flash while infants were fixated on the light. As in the familiarization phase, if the infant continued to look at the light for 20 seconds, the trial ended automatically and the next trial began. Similarly, if the infant failed to look at the side light for at least 2 seconds, the trial was automatically repeated. A minimum criterion of 2 seconds was necessary to allow the infant to hear at least one token of the target word in a sentence.

Results and Discussion

In this experiment, infants were exposed to two words. One word was presented at high amplitude and one at lower amplitude. Therefore, there were three dependent measures of interest: listening times to the sentences containing the amplitude-matched word, listening times to the sentences containing the amplitude-mismatched word, and listening times to the sentences containing unfamiliar words (averaged across both types of unfamiliar words). A one-way ANOVA (match/mismatch/unfamiliar) revealed a marginally significant main effect of stimulus type $F(2, 45) = 3.003$, $p=.06$. Individual comparisons revealed that for sentences containing matched words, infants showed significantly higher listening times compared with those containing unfamiliar words,

$t(15) = 2.91$, $p<.05$. Twelve of sixteen infants showed this pattern. For mismatched words, infants also showed significantly higher listening times for passages containing mismatched words compared with those containing unfamiliar words, $t(15) = 2.61$, $p<.05$. Ten of sixteen infants showed this pattern. A comparison of listening times revealed no significant difference in listening times to sentences containing matched and mismatched items, $t(15) = .53$, NS. Listening times are depicted in Figure 1.

Insert Figure 1 about here

These findings demonstrate that infants are able to recognize words in fluent speech, even if those words differ in amplitude, raising striking parallels with findings in the adult lexical processing literature. Like infants, adults are unimpeded by changes in amplitude across tokens (in either speed or accuracy of making lexical decisions). In this respect, infants and adults appear to be very similar in the types of surface details they are predisposed to disregard.

The findings of this study call into question a simple covariance account of early word recognition, which would predict that prelexical infants only recognize words that match in surface form across encounters. In this case, each familiarization item was repeated in either high or low amplitude, so that a particular loudness level was reliably associated with a particular word. In spite of this, infants were able to recognize repetitions of words across significant changes in amplitude. This contrasts with previous findings on talker gender, speaker affect, and emphatic stress (Bortfeld & Morgan, in submission; Houston & Jusczyk, 2000; Singh et al. 2004), which showed that changes in surface form disrupt word recognition in infants at this age.

This set of findings leaves open the question of why amplitude is treated in a manner distinct from vocal affect and talker gender, where changes in the former do not disrupt spoken word recognition but changes in the latter do. A relatively straightforward hypothesis, introduced in discussions of adult speech processing (Sommers et al., 1994), is that the *complexity* of acoustic changes introduced by different types of surface variation mediates speed and accuracy of processing. This explanation is based on the assumption that generalizing across tokens that differ in complex ways is resource-consuming. Changes in gender and affect involve simultaneous, complex, non-linear changes at multiple levels of the acoustic signal (e.g. F0, speech rate, voice quality). As a result, deriving the relationship between a token produced in a happy voice and another produced in a neutral voice, for example, is difficult. By contrast, if amplitude has been altered uniformly across time intervals, deriving a low amplitude token from a high amplitude token is much more straightforward, involving simple scaling of one dimension of the signal by a constant value. Thus, it is possible that infants' word recognition is only impaired when they must simultaneously track variation across tokens along numerous acoustic dimensions. Infants' failure to recognize words across changes in talker and affect, and (as we have shown in Study 1) their successful recognition of words across changes in amplitude are findings consistent with this 'complexity' hypothesis, as both talker and affect have a relatively complex set of acoustic correlates but amplitude does not.

However, before concluding that the complexity hypothesis is correct, it is necessary to explore another simple acoustic dimension, e.g., pitch. The complexity hypothesis would predict that infants should be able to ignore variations in pitch as well as they can ignore amplitude since both are single dimensions. That is, that infants will recognize tokens that are scaled along a single acoustic dimension, whether it is amplitude or pitch. In the event that infants succeed in word recognition with one type of variation (e.g. amplitude), but fail to recognize words when presented with a different type of variation (e.g. pitch), alternative hypotheses will have to be formulated to account for a selective mechanism that allows recognition across certain surface changes but not others. The next experiment assesses word recognition amidst simple changes in pitch to determine whether infants are able to tolerate such changes not only in amplitude but also in another single acoustic dimension.

EXPERIMENT 2

In this study, 7.5-month-old infants were familiarized with two words, one in a relatively high pitch and another in a relatively low pitch. Infants then heard passages with sentences containing the familiarized words, as well as comparable passages containing nonfamiliarized words.

<u>Participants</u> Thirty-two, English-exposed 7.5-month-olds participated in the study (12 males and 20 females), recruited from Massachusetts birth records and advertisements. Mean age of participants was 224 days (range = 211 days to 247 days). Data from two additional infants were not included for inattention.

<u>Stimuli</u> The ‘happy’ stimulus set used in Singh et al. (2004) was modified to create two new stimulus sets, each involving a uniform transformation of the fundamental frequency. Fundamental frequency is the primary acoustic correlate of perceptual changes in pitch. One set of stimuli (henceforth the High Pitch set) was created by raising the fundamental frequency of all words and passages by ¼ octave (3 semitones). This

was done by applying a uniform translation of all pitch points up by ¼ octave. A second set of stimuli (henceforth the Low Pitch set) was created by decreasing the fundamental frequency of all words by the same amount (¼ octave). Therefore, the difference between the two sets of stimuli was half an octave (6 semitones). Both sets of stimuli involved pitch manipulations, so that infants' preferences would not be affected by the naturalness of the stimuli. These particular values were chosen because they are comparable to the more complex pitch changes (i.e. due to affect) introduced in previous studies (Singh, et al., 2004) and are well within the natural range of pitch values expressed in infant-directed speech (Garnica, 1977). All acoustic transformations and measurements were performed using a script in the Praat program (Boersma & Weenink, 1996). Amplitude and duration measures were identical across the two stimulus sets. Pitch measurements are shown in Table 1.

During familiarization, infants heard citation form tokens of two words. Half of the infants heard the words *bike* and *hat* while the other half heard *tree* and *pear*. For each infant, one word was heard in High Pitch and the other in Low Pitch. Recognition passages were either all in High Pitch or all in Low Pitch, matched with one familiarization item and mismatched with the other. During recognition testing, infants heard passages containing all four words. As a result of counterbalancing, familiar passages for some infants were unfamiliar to others and vice versa.

Results and Discussion

Infants in this experiment were familiarized with two words, one in a high pitch and one at a lower pitch. Recognition passages, some of which contained the familiarized words (familiar passages) and some of which did not (unfamiliar passages), were either presented in a high or low pitch. A 2X2X2 ANOVA (match status by familiarity by passage type) revealed an interaction of match status and familiarity, $F(1,60)=6.85$, $p<.05$ and a significant main effect of familiarity, $F(1,60)=17.87$, $p<.001$. There was no significant interaction of passage type with match status or familiarity. Planned comparisons revealed significantly higher listening times to passages containing matched (high) familiarization items in high passages compared with unfamiliar items, $t(15) = 4.59$, $p<.001$. Thirteen of sixteen infants showed this pattern. By contrast, listening times for passages containing mismatched items (low familiarization words presented in high sentences) were not significantly different from listening times to unfamiliar passages, $t(15) = 1.2$, NS. Similarly, infants presented with low recognition passages demonstrated significantly higher listening times to passages containing matched (low) familiarization items, $t(15) = 3.5$, $p<.05$ and listening times to mismatched words that did not depart significantly from listening times to unfamiliar passages, $t(15) = .32$, NS. Twelve of sixteen infants showed higher listening times for matched items than for unfamiliar passages. There was no significant effect of passage type (high or low). Listening times to each type of stimulus during the recognition phase are depicted in Figure 2.

Insert Figure 2 about here

Infants therefore recognized only words that were *matched* in pitch across familiarization and recognition phases at 7.5 months. As with talker gender and speaker affect, infants failed to recognize words that were mismatched in pitch across tokens. This suggests that infants at this age use pitch information to classify words, even though it is not lexically relevant in English. Therefore, infants show different capacities to cope with pitch variation and amplitude variation across tokens; their performance is disrupted by the former and unaffected by the latter. This presents a challenge to the complexity hypothesis, which assumes that infants can disregard simple changes across words and are affected only by complex changes across tokens. Both pitch and loudness changes, in the way in which they were generated in this set of studies, involved very simple changes to one acoustic dimension of the signal. In spite of this simplicity, infants showed the capacity to disregard amplitude changes but not pitch changes. Infants' inability to cope with pitch changes in early word recognition more closely resembles their performance on tasks where talker gender and vocal affect are manipulated. In sum, it appears that the ease of normalization is not equal for different dimensions of surface change.

Before considering alternative accounts, it is important to find out when infants begin to succeed at recognizing word tokens that differ in pitch. At a minimum, this process involves differentiating phonological cues that communicate lexical information in the native language from those that do not. Evidence that infants have begun to appreciate acoustic properties of consonants of their language appears by approximately 8 to 9 months. After this age, infants begin to map consonant sounds onto phonetic categories with reference to the structure imposed by their native language (Best, 1995; Werker & Tees, 1983), they begin to appreciate the phonotactic regularities inherent in their language (Mattys & Juszcyk, 2001), and they have mastered the stress pattern of their native language (Echols, Crowhurst & Childers, 1997). It is plausible that infants also begin to disregard lexically irrelevant detail in spoken word recognition during this period (with the possible exception of amplitude, which is disregarded prior to 7.5 months). Indeed, one aspect of native language attunement is arriving at the phonetic determinants of lexical relevance (i.e. phonemic properties) in the native language.

Therefore, in a third experiment, we investigated infants' abilities to overcome pitch variation in spoken word recognition later in development, at 9 months.

EXPERIMENT 3

In this experiment, a group of 9-month-old infants was familiarized with two words, one in a relatively high pitch and the other in a relatively low pitch. Given that there was no effect of the pitch of recognition passages on listening times to familiarized stimuli, infants in this experiment heard all recognition passages spoken in a high pitch only.

Participants Sixteen full-term, English-exposed 9-month-olds participated in the study (9 males and 7 females), recruited from Massachusetts birth records and advertisements. The mean age of participants was 276 days (range = 265 days to 297 days).

Stimuli, apparatus and procedure Stimuli, apparatus and procedure were identical to those in Experiment 2.

Results and Discussion

As in the previous experiment, infants were familiarized with two words, one in a matched pitch (similar pitch levels across familiarization and recognition) and one in a mismatched pitch (dissimilar pitch levels across familiarization and recognition). A one-way ANOVA revealed a significant main effect of stimulus type (matched/mismatched/unfamiliar), $F(2,45) = 8.65$, p<01. Individual comparisons revealed that for the matched familiarization words, infants showed significantly higher listening times to passage containing these words compared with unfamiliar passages, $t(15) = 3.95$, $p<.01$. Fourteen of sixteen infants showed this pattern of results. Unlike infants at 7.5 months, however, in this study, infants also showed significantly higher listening times to mismatched words compared with passages containing unfamiliar words, $t(15) = 2.39$, $p<.05$. Twelve of sixteen infants showed this pattern of results. A comparison of listening times to sentences containing matched and mismatched words revealed no significant difference, $t(15) = 1.38$, NS. Listening times are depicted in Figure 3 to each type of sentence.

While a 2X3 ANOVA (age by stimulus type) that compared data across Experiments 2 and 3 (Figures 2 and 3) did not reveal a significant interaction, individual planned comparisons revealed distinct patterns in how matched and mismatched items were treated by older and younger infants. This analysis was only done on the condition where recognition passages were presented in a high pitch because both older and younger infants were tested in that condition. There was no significant difference in listening times to matched and mismatched stimuli for older infants, $t(15) = 1.38$, NS, and a marginally significant difference in listening times to these two types of stimulus in younger infants, $t(15) = 1.84$, $p=.08$, with infants listening more to matched stimuli than to mismatched stimuli. This suggests that while younger infants listened differentially to matched and mis-matched words, older infants did not, evidenced by their capacity to recognize both types of familiarized words.

Insert Figure 3 about here

Together, these findings demonstrate that at the earliest stages of spoken word recognition, infants are highly sensitive to at least some aspects of the physical form in which a word is encountered. However, the present results suggest that by 9 months, infants are able to reliably recognize words across changes in pitch. While it is not clear specifically what information infants garner about words between 7.5 and 9 months, this developmental period also marks important shifts in infants' progression to a native-language perceptual frame of reference for other components of the language code (Best, 1995). Therefore, it is perhaps not surprising that infants also demonstrate greater knowledge about the determinants of lexical relevance in their native language at this point in development.

GENERAL DISCUSSION

In this set of studies, we sought to determine how infants cope with different types of non-phonemic variability in the speech stream. Infants at 7.5 months were tested on their capacities to detect recurrences of words when those words differed in pitch and in amplitude across tokens. Our findings showed that at this early stage of development, infants were unable to recognize words unless they were similar in pitch across familiarization and recognition phases of the experiment. However, they could generalize across variation in amplitude. By 9 months, infants were not only able to generalize to new instances of words that differed in amplitude evidenced at 7.5 months, but were also able to generalize to new instances of words that differed in pitch, displaying more adult-like word recognition abilities.

Our results add to a growing body of evidence showing that when infants begin to segment and track words from fluent speech at around 7.5 months, their lexical representations incorporate non-phonemic as well as phonemic characteristics of words. Previous studies have shown this to be true with respect to talker identity (Houston & Jusczyk, 2000), speaker affect (Singh et al., 2004), and emphatic stress (Bortfeld & Morgan, in submission). In each case, change in the variable of interest appears to disrupt infants' recognition of familiarized words. Here, we have shown that infants are even disrupted by changes involving simple scaling of *one dimension* of the signal, in this case, pitch. However, they are not disrupted by all changes involving a single dimension, as they are able to recognize words presented at different amplitude levels.

This set of findings provides clues to the structure of the early lexicon. That infants retain at least some episodic properties of word tokens they encounter suggests that their lexical representations comprise exemplars, rather than normalized abstractions. Storage of exemplars, though perhaps costly in memory, offers an escape from what would otherwise be an insoluble bootstrapping problem. Languages of the world differ in the dimensions of sound they exploit to convey meaning, so sound variation that constitutes mere surface detail in one language may be lexically relevant in another. Infants must discover which dimensions of sound are relevant to meaning in order to reliably recognize words in fluent speech; they cannot know ahead of time which dimensions these will be for the language they are learning. Infants cannot construct abstract lexical representations without knowledge of which dimensions are important for lexical contrast. If infants store exemplars, as several models have suggested (Juszcyk, 1993, 1997; Werker & Curtin, 2005), surface variations that are orthogonal to lexical identity, will be washed out by the aggregate activation of stored traces. Moreover, across time, infants can learn how to weight various aspects of the signal (Jusczyk, 1993; 1997). At 7.5 months, infants apparently encode most dimensions of sound with the expectation that they may be lexically relevant. Before the end of the first year, however, infants have discovered that changes in pitch, vocal affect, and talker identity are not relevant to lexical identity perhaps because they have noted that these dimensions vary orthogonally to invariant phonemic detail. Treatment of amplitude appears to be exceptional and not subject to the same developmental influences, evidenced by the finding that infants disregard amplitude earlier in development than pitch, vocal affect, and talker identity.

Traditional models of the adult mental lexicon assumed that all surface details that are not lexically relevant are stripped away via normalization processes (Blumstein & Stevens, 1979; Gerstman, 1968; Shankweiler, Strange & Verbrugge, 1977; Stevens, 1972). More recently, a considerable amount of research probing the architecture of the mental lexicon has revealed surprising precision and detail in adults' memories for words (e.g. Church & Schacter, 1994; Goldinger, 1996; 1998; Mullenix, Johnson, Topcu-Durgun & Farnsworth, 1995; Palmeri, Goldinger & Pisoni, 1993). This research has supported the view that adults encode individual exemplars of words rather than prototypes of words in memory (e.g., Goldinger, 1998; Tenpenny, 1995). On this view, the degree of similarity between newly encountered tokens and previous tokens of a given word mediate the speed and accuracy with which new tokens will be processed. The present study does not support a strict interpretation of either view of spoken word recognition. While infants are clearly aided by surface matches, they do appear to be able to normalize for some surface properties (i.e. amplitude) at an early age.

Given the ephemeral nature of the signal, spoken word recognition must operate under a fairly strict deadline. If infants' lexical processing is initially inefficient, as recent findings suggest (e.g., Fernald, Pinto, Swingley, Weinberg, & McRoberts, 1998), then superficial variations that merely slow processing in adults may well thwart infants' word recognition. This is consistent with the notion that there is continuity in the architecture of lexical processing mechanisms between infancy and adulthood, though there may be quantitative changes across development due to increased efficiency of processing. In the present set of studies, we have simply measured the accuracy with which infants recognize words. In adult psycholinguistics, it is possible to record both the speed and accuracy with which adults recognize words amidst surface variation. Typically in adults, disruptive effects of surface variation manifest themselves in the speed of lexical access, rather than accuracy. Therefore, our findings are not directly comparable to those generated with adults, given the inevitable differences in experimental procedures. It is possible, however, that if we were able to measure the speed of lexical access in infants, older infants would demonstrate reduced speed of access in response to pitch variation, although perhaps not in response to amplitude variation. Such findings would point to similarity between older infants and adults both in the efficiency and accuracy of lexical access in the face of different types of surface variation.

Also consistent with the claim of architectural continuity is our finding that certain types of surface changes do *not* impede early word recognition. Importantly, infants' and adults' word recognition abilities are impervious to variation in amplitude (Bradlow, Nygaard & Pisoni, 1999; Church & Schacter, 1994; Nygaard, Sommers & Pisoni, 1995). Although both adults and infants are sensitive to changes in talker, speaker affect and pitch, neither population appears to be affected by amplitude changes across tokens. The fact that amplitude is not, in and of itself, lexically relevant in any language leads to possible hypotheses about why it is disregarded in spoken word recognition.

The current set of findings does not conform to a complexity account nor to a covariance account. Infants appear to be able to tolerate relatively simple changes in amplitude, but not in pitch, suggesting that transformations along a single dimension do not affect word recognition uniformly, as required by the complexity account. Similarly, infants are not able to recognize words mismatched in pitch yet are able to recognize words mismatched in amplitude, calling into question a covariance account. Why is amplitude exceptional? One possibility is that infants in particular might have difficulty calibrating amplitude, so that they fail to detect changes along this dimension. Sinnott & Aslin (1985), however, have shown that 7-9-month-olds discriminate intensity changes, so this cannot be the correct explanation. Alternatively, by 7 months or so, infants may have accumulated sufficient exposure to words spoken with varying amplitudes to have learned that amplitude is not germane to lexical identity. While infants are also likely to have had considerable experience with pitch changes, pitch changes as an optional part of speech controlled by the speaker. Amplitude changes, by contrast, can be controlled by the speaker, but also by the listening environment. For example, barriers that may partially mask sound between speakers and listeners as well as distance between speakers and listeners impact upon amplitude levels but not upon properties such as pitch. Therefore, infants are likely to have heard familiar words uttered by speakers who are nearer (hence louder) or farther away (hence softer). Moreover, in addition to showing greater frequency modulation than adult-directed speech, child-directed speech also shows greater amplitude modulation: several researchers have noted that whispering is common in child-directed speech (e.g., Fernald & Simon, 1984). While both pitch and amplitude are modulated in infant-directed speech, for the aforementioned reasons, infants' experiences with amplitude modulation may be more extensive and varied by 7.5 months. Therefore, it is possible that early and constant experience with amplitude variation may lead infants to judge amplitude variation alone as irrelevant for classifying linguistic stimuli. An analogous developmental trend emerges in the visual domain. For example, it has been shown that visual experience helps infants determine how visual changes reflect categorical changes in color. Early in development, infants easily disregard certain types of visual changes, such as luminance, presumably because repeated exposure to changes in luminance from birth reveals that objects of the same color can appear different under different lighting conditions. As a result, infants do not take luminance variation into account when they judge how to partition the color spectrum, but do attend to other changes that are comparable in complexity, such as hue and saturation (Kaldy, Blaser & Leslie, 2004).

If infants *learn* to ignore amplitude variation, then they ought to be able to learn to (re)attend to such variation, at least in specific contexts. To test this, we are currently conducting a study in which 7.5-month-olds are first presented with a series of words, each spoken with both high and low amplitude. When a word occurs with high amplitude, it is paired with one visual stimulus; when it occurs with low amplitude, it is paired with another. Infants are then familiarized with two new words and tested as in Experiment 1. If infants can learn that amplitude does affect lexical identity in this context, they should behave like the infants in Experiment 2, recognizing only those words that match across familiarization and test.

On the other hand, amplitude may be exceptional, not because infants have learned that amplitude variation is irrelevant for classifying speech stimuli, but because of the special status of amplitude in specifying the location of a source. As noted, amplitude tends to vary as a function of the distance between speaker and hearer. Of course, this relation is not specific to linguistic stimuli; in general, amplitude serves as an important cue to the proximity of any audio source (Bregman, 1990). It is not difficult to conjecture that built-in knowledge of this amplitude-proximity relation would be useful to a very wide range of species. If infants are geared to use amplitude level as a means of identifying audio source location, perhaps this pre-empts use of this cue for lexical identification. More generally, perhaps auditory scene analysis takes precedence over auditory object/event identification, so that *any* cue functioning indexically in a given context is no longer available for use in identification. The fact that amplitude, in and of itself, is not exploited to communicate meaning in human languages lends plausibility to this hypothesis.

In summary, the present set of studies attests to infants (like adults) having a mental lexicon that stores words in a way that selectively preserves pitch detail in a way that proves detrimental to recognizing words that contrast in pitch until 9 months. While infants do encode non-phonemic surface detail perhaps by virtue of not yet having access to meaning, they do not do so indiscriminately. Young infants appear to treat pitch, but not amplitude, as lexically relevant. While both of these changes represent relatively simple transformations in the signal in the present studies, compared to more complex transformations of vocal affect and talker gender that have been investigated thus far, infants treated the two factors differently. Whether the asymmetry observed between amplitude and other surface properties results from infants' experience with words or from innate predispositions remains unclear. Nevertheless, the current set of findings establishes a developmental trajectory by which infants progress to more mature processing of words and provides a clearer understanding of how infants treat particular surface details when they encode words in memory.

References

Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology, 70,* 614–636.

Best, C. T. (1995). Learning to perceive the sound patterns of English. In C. Rovee-Collier & L. P. Lipsitt (Eds.), *Advances in infancy research* (pp. 217–304). Norwood, NJ: Ablex.

Blumstein, S.E. and K.N. Stevens (1979). Acoustic invariance in speech production:Evidence from measurements of the spectral characteristics of stop consonants. *Journal of the Acoustical Society of America.* 66: 1001-1017.

Boersma P.& Weenick D.J.M, 1996. Praat, a system for doing phonetics by computer, *Institute of Phonetics Sciences of Amsterdam, Report 132*.

Bortfeld, H., & Morgan, J. L.(in submission). Effects of focusing stress on infants' fluent speech word recognition.

Bortfeld, H., Morgan, J.L., Golinkoff, R.M. & Rathbun, K. (2005). Mommy and me: Familiar names help launch babies into speech stream segmentation. *Psychological Science, 16*, 298-304.

Bradlow, A. R., Nygaard, L. C., & Pisoni, D. B. (1999). Effects of talker, rate, and amplitude variation on recognition memory for spoken words. *Perception & Psychophysics*, *61*, 206–219.

Bregman, A. S. 1990. *Auditory Scene Analysis*. Cambridge, MA: MIT Press.

Church, B. A., & Schacter, D. L. (1994). Perceptual specificity of auditory priming: Implicit memory for voice intonation and fundamental frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 521–533.

Echols, C. H., Crowhurst, M.J., & Childers, J. B. (1997). The perception of rhythmic units in speech by infants and adults. *Journal of Memory and Language*, *36,* 202-225.

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

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

Garnica, O.K. (1977) Some prosodic and paralinguistic features of speech to young children. In C. E. Snow & C. A. Ferguson (Eds.), *Talking to children: Language input and acquisition* (pp. 63-88). Cambridge: Cambridge University Press.

Gerstman, L. J.(1968). Classification of self-normalized vowels, *IEEE Transactions on Audio Electroacoustics*, *AU-16*, 78-80.

Goldinger, S. D. (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22,* 1166–1183.

Goldinger, S. D. (1998). Echoes of echoes. An episodic theory of lexical access. *Psychological Review*, *105*, 251–279.

Houston, D.M. (2000). *The role of talker variability in infant word representations.* (Doctoral Dissertation, Johns Hopkins University, 1999). Dissertation Abstracts International, 60, 5802.

Houston, D. M., & Jusczyk, P. W. (2000). The role of talker specific information in word segmentation by infants. *Journal of Experimental Psychology: Human Perception and Performance, 26,* 1570–1582.

Jusczyk, P. (1993). From general to language-specific capacities: The WRAPSA model of how speech perception develops. *Journal of Phonetics, 21,* 3–28.

Jusczyk, P. W. (1997*). The discovery of spoken language*. Cambridge, MA: MIT Press.

Jusczyk, P. W., & Aslin, R. N. (1995). Infants detection of the sound patterns of words in fluent speech. *Cognitive Psychology, 29,* 1–23.

Kaldy, Z., Blaser, E., & Leslie, A. (2004). *Iso-salient color and luminance information in object identification*. Poster presented at the International Conference on Infant Studies, May 5-8, Chicago, IL.

Kemler Nelson, D. G., Jusczyk, P. W., Mandel, D. R., Myers, J., Turk, A., & Gerken, L. A. (1995). The headturn preference procedure for testing auditory perception*. Infant Behavior and Development, 18,* 111–116.

Luce, P. A., & Lyons, E. A. (1998). Specificity of memory representations for spoken words. *Memory & Cognition, 26,* 708–715.

Mattys, S. L., & Jusczyk, P. W. (2001). Phonotactic cues for segmentation of fluent speech by infants. *Cognition, 78,* 91–121.

Mertus, J. (2002). *Brown Lab Interactive Speech System*. Brown University.

Miller, J. L., & Lieberman, A. M. (1979). Some effects of later occurring information on the perception of stop consonants and semi-vowels. *Perception and Psychophysics, 25*, 457-465.

Mullenix, J. W., Johnson, K. A., Topcu-Durgun, M., & Farnsworth, L. W. (1995). The perceptual representation of voice gender. *Journal of the Acoustical Society of America, 98*(6), 3080-3095.

Murray, I. R., & Arnott, J. L. (1993). Toward the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion. *Journal of the Acoustical Society of America, 93,* 1097–1108.

Nygaard, L. C., Sommers, M. S., & Pisoni, D. B. (1995). Effects of stimulus variability on perception and representation of spoken words in memory. *Perception & Psychophysics, 57*, 989–1001.

Palmeri, T. J., Goldinger, S. D., & Pisoni, D. B. (1993). Episodic encoding of voice attributes and recognition memory for spoken words. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 309–328.

Pisoni, D. B. (1997). Some thoughts on "Normalization" in speech perception. In K. Johnson & J. W. Mullennix (Eds.), *Talker Variability in Speech Processing*. San Diego: Academic Press, pp. 9-32.

Remez, R. E., Fellowes, J. M., & Rubin, P. E. (1997). Talker identification based on phonetic information. *Journal of Experimental Psychology: Human Perception and Performance*, 23, 651-666.

Scherer, K. (1986). Vocal affect expression: A review and a model for future research. *Psychological Bulletin, 9*, 143–165.

Shankweiler, D., Strange, W. & Verbrugge, R. (1977). Speech and the problem of perceptual constancy. In R. Shaw, & J. Bransford (Eds*.), Perceiving, acting, and* *knowing: Toward an ecological psychology*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Singh, L., Morgan, J., White, K. (2004). Preference and processing: The role of speech affect in early spoken word recognition. *Journal of Memory and Language*, *51*(2), 173-189.

Sinnott, J. M., and Aslin, R. N. (1985). Frequency and intensity discrimination in human infants and adults. *Journal of the Acoustical Society of America, 78,* 1986–1992.

Sommers, M.S., Nygaard, L.C. & Pisoni, D.B. (1994). Stimulus variability and spoken word recognition: Effects of variability in speaking rate and overall amplitude. *Journal of the Acoustical Society of America, 96,* 1314-1324.

Stevens, K.N. (1972) "The quantal nature of speech:Evidence from articulatory-acoustic data." In P.B. Denes and E.E. David Jr. (eds.), *Human Communication, A Unified View* New York:McGraw-Hill, pp. 51-66.

Tenpenny, P. L. 1995. Abstractionist versus episodic theories of repetition priming and word identification. *Psychonomic Bulletin & Review 2*, 339-363.

Werker, J.F. & Curtin, S. (2005). PRIMIR: A Developmental Framework of Infant Speech Processing. *Language Learning and Development*, *1*(2), 197-234.

Werker, J. F., & Tees, R. C. (1983). Developmental changes across childhood in the perception of non-native speech sounds. *Canadian Journal of Psychology*, *37*, 278–286.

Williams, C. E., & Stevens, K. N. (1972). Emotions and speech: Some acoustical correlates. *Journal of the Acoustical Society of America, 52*, 233–248.

Table 1

Acoustic Analyses of Words and Sentences

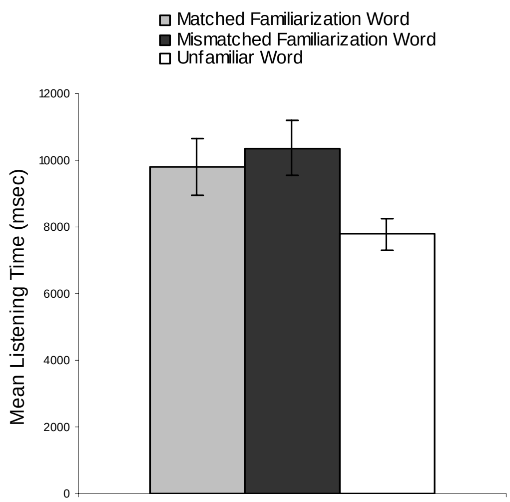
	Mean Fundamental Frequency (Hz) (SD)		Mean Amplitude (dB) (SD)	
	Words	Sentences	Words	Sentences
Experiment 1: Low Amplitude	-	-	57.11 (3.02)	
Experiment 1: High Amplitude	-	-	76.76 (3.82)	73.05 (1.47)
Experiment 2: Low Pitch	269.38 (43.91)	240.7183 (19.48)	-	-
Experiment 2: High Pitch	383.718 (61.46)	339.1065 (30.22)	-	-

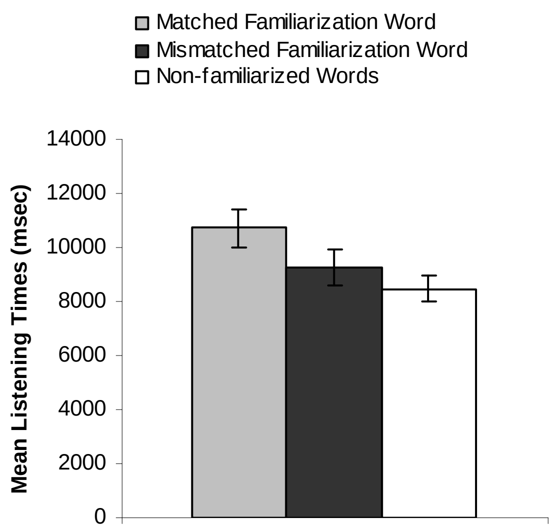
Figure Captions

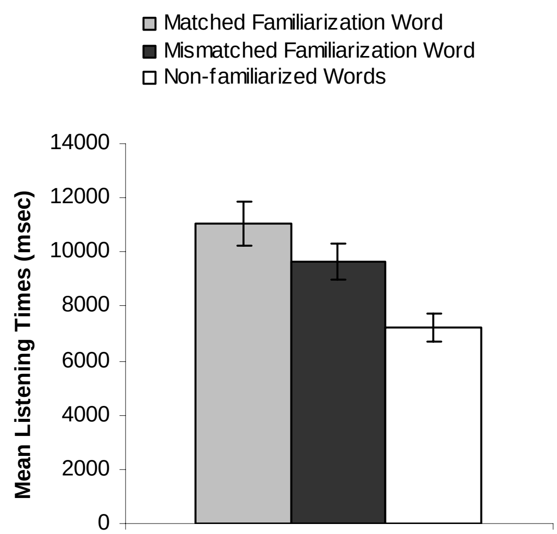
Figure 1 Experiment 1: 7.5-month-olds' listening times to matched, mismatched and unfamiliar stimuli with standard errors for high and low amplitude words.

Figure 2 Experiment 2: 7.5-month-olds' listening times to matched, mismatched and unfamiliar stimuli with standard errors for high and low pitched familiarization words.

Figure 3 Experiment 3: 9-month-olds' matched, mismatched and unfamiliar stimuli with standard errors for high and low pitched familiarization words.


Matched Familiarization Word
Mismatched Familiarization Word
Unfamiliar Word
Mean Listening Time (msec)
12000
10000
8000
6000
4000
2000
0


Matched Familiarization Word
Mismatched Familiarization Word
Non-familiarized Words
Mean Listening Times (msec)
14000
12000
10000
8000
6000
4000
2000
0


Matched Familiarization Word
Mismatched Familiarization Word
Non-familiarized Words
Mean Listening Times (msec)
14000
12000
10000
8000
6000
4000
2000
0

Appendix

Words and passages used in recognition testing

Bike
His bike had big black wheels
The girl rode her big bike
Her bike could go very fast
The bell on the bike was really loud
The boy had a new red bike
Your bike always stays in the garage.

Hat
She put on her hat to play in the snow.
The hat was soft and warm
Her brother had knitted the hat
The hat was blue and white
She liked how the hat covered her ears
Her friends also liked her hat.

Tree
The tree was a hundred years old
The tree grew in the man's back yard
He liked to look outside at the tree
Hanging from the tree was a swing
The man's grandchild played in the tree.
The leaves on the tree were yellow

Pear
The juicy, green pear came from the basket
The pear is her favorite fruit.
She wanted to eat the biggest pear.
The pear in the basket looked very good.
Next to the pear was an apple.
She ate the whole pear.