Part I

Introduction to Speech Perception

Introduction to Part I

Speech perception and its development in children is one of the richest and liveliest subfields in language research. Because speech is a physical as well as a psychological phenomenon, it has lent itself more easily to experimental study than some of the "higher" (and more controversial) levels of language and language processing. This solid grounding in reality has permitted a remarkable amount of progress – and yielded a lot of big surprises.

In the past few decades a series of technical breakthroughs have made it possible for scientists to visualize the acoustic signals that support speech, using sound spectrograms and other instruments. Figure 1 (below, from Aslin et al., 1983) illustrates a schematic ("cleaned up") version of the sound spectrogram for two syllables, \di\ (pronounced "dee") and \gu\ (pronounced "goo"). Time in milliseconds is plotted on the horizontal axis. The vertical axis represents a graded series of frequency bands, or "formants" (as in a car radio tuner). The black bands in this two-dimensional space represent changes over time in the distribution of energy within those frequency bands. The visible shifts in Figure 1.1 of energy from one formant to another are called "formant transitions," long believed to play a key role in signaling systematic contrasts between speech sounds (e.g., between "d" and "g"). With such instruments, scientists are able to modify the signal in many different ways, and then play it back to see how those modifications sound to real human beings. A major motivation for this research (then and now) has been to construct artificial systems that can understand speech, so we can talk to our computers directly, bypassing the keyboard. In principle, this kind of mechanism might also be very useful to people who are congenitally deaf.

Considering all the time and money that has gone into this enterprise, the first big surprise is that we still have so far to go. Today's computers can be trained to understand a finite set of words uttered by a single speaker ("his master's voice"), or a very small set of words uttered by a much larger array of speakers (like the numbers 0–9, or the words "calling card" and "collect" that we find ourselves yelling at telephone company computers). Much more flexible and sophisticated speech-understanding systems may be just around the corner. But we are still well below the level reached by healthy human infants across the first year of life. The problem of speech perception has proven to be especially difficult to solve, because the relationship between physics (the actual acoustic events, like those in Figure 1) and experience (the sounds we hear) is not at all transparent. Three examples of this problem include the following.

- Violations of linearity. If the mapping from sound to experience were straightforward, then we should expect the first part of each pattern in Figure 1 to sound

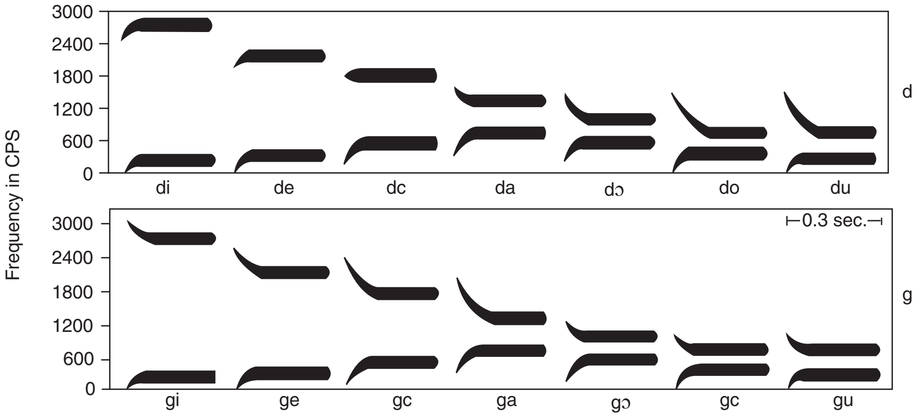

Figure 1.1

like a consonant ("d" or "g") while the second part would sound like a vowel ("eee" or "ooo"). For the vowel portion, the prediction works. However, when the formant transitions that signal different consonants are played back without their vowel contexts, they do not sound like speech at all! Instead, they sound like clicks or brief chirps that disappear when these bits of sound are placed back into a speech context.

- Violations of invariance. If the mapping from sound to experience were transparent, then we would also expect the formant transitions that signal "d" and "g" in Figure 1 to play more or less the same role in other vowel contexts. So if we spliced the first part of \di\ in front of the vowel \u\, then we would expect the resulting spectrogram to sound like \du\. But that is not what happens! These bits of sound do not behave like letters of the alphabet. For example, the "d" component of the syllable "du" looks like the "g" component of the syllable "ga." Furthermore, the shape of the visual pattern that corresponds to a constant sound can even vary with the pitch of the speaker's voice, so that the "da" produced by a small child results in a very different-looking pattern from the "da" produced by a mature adult male. The bottom line is that the physical components that make up speech sounds are not invariant over contexts; they are highly context dependent, changing their colors completely depending not only on the vowels they precede and follow, but on the voice of the person who is doing the talking (i.e., the "fundamental" or carrier frequency that characterizes the difference in sound quality between men and women, or adults and children). We perceive speech sounds as "same" or "different," but the basis for this difference in experience is not obvious from the physical signal itself.
- Categorical perception. Consonants like \p\ and \b\ differ along a dimension called voice onset time (VOT), a difference between the point in time at which the vocal chords begin to vibrate and the discontinuous point at which we open our lips to allow that continuous sound to emerge. It is possible to make up artificial tokens of \p\ and \b\ that differ continuously along this VOT dimension. However, native speakers do not hear this as a gradual change. Instead, they hear a sudden or "categorical" transition from \p\ to \b\. The physical basis for this

categorical shift is not at all obvious, a surprising result that led some investigators to conclude that the boundary is imposed by the listener's auditory processing system.

As peculiarities like these began to mount, it became obvious that speech-understanding systems would not be as easy to construct as we had originally hoped. Some investigators concluded that human beings process speech the way they do because we have a special-purpose speech perception device built into our brains, imposing psychological experiences upon an underdetermined physical event. Some of the pioneers of speech perception research also proposed that this innate, special-purpose device is based not on audition per se but on the human system for speech production. This theory (called the Motor Theory of Speech Perception) is based on a kind of "analysis by synthesis": that is, we perceive auditory input as speech by "coming up to meet it" with an internal model of what that person we are listening to was trying to produce with an articulatory system very much like our own. This theory led to several clear predictions: (1) speech perception should be unique to humans; (2) speech perception makes use of a neural substrate that is separate from the neural system used for other kinds of audition; (3) the system should be innate, up and running at birth.

Evidence in favor of this view (especially the third point) began to appear in 1975, with an influential paper by Peter Eimas showing that very young infants (2–4 months old) are not only able to hear the phoneme contrasts that characterize natural languages, but to hear them categorically (Eimas, 1975). Habituated on one set of sounds (e.g., \ba\), the infants showed signs of surprise (e.g., vigorous sucking on an electronically monitored pacifier) when the signal shifted to \pa\, with a sharp border roughly around the point at which adults also show a categorical boundary. Following this discovery, many more studies of infant speech perception, using a variety of methods, led to the clear conclusion that infants are born able to perceive most if not all of the speech contrasts used by natural languages. As Patricia Kuhl has put it, infants are born "citizens of the world," able to hear and learn any natural language without prejudice.

It is now quite clear that the ability to perceive speech contrasts is present very early, and is probably (within limits) an innate property of the human auditory system. This does not mean, however, that our innate perceptual abilities are unique to speech, or unique to humans. For example, subsequent studies have shown that categorical perception also occurs with sequences of pure tones (Cutting & Rosner, 1974) and with sequences of lights (the famous flicker-fusion phenomenon). Perhaps the most important and surprising finding in this regard lies in a growing literature showing that categorical perception of speech sounds occurs in non-human species, e.g., chinchillas (Kuhl & Miller, 1975) and quail (Lotto et al., 1997). The article by Ramus et al. (2000, and in this volume) pushes these observations one step further, showing that tamarin monkeys can discriminate the same speech contrasts perceived by human infants, and are (like human infants) unable to make those discriminations when these speech sequences are played backwards – suggesting that the pereception is not just of random sound sequences but rather of something concerning patterned human speech.

The overwhelming conclusion from studies like these is that human speech perception evolved to exploit pre-existing dimensions and categories that were already present in the mammalian auditory system. Crudely put, the mouth evolved to meet the ears, and not vice versa. Given such evidence, research on speech perception and its development has shifted from the initial state of the organism (What can humans perceive at birth? Virtually all speech contrasts – but this is not unique to humans) to a focus on the process by which children learn to tune in to the 40 or so phonemes used by their own native language (out of an array of up to 4,000 possibilities). Even more surprises have emerged out of this research effort. For example, we now know that human infants develop a bias toward certain sounds from their native language in utero (DeCasper et al., 1994; Ramus et al., 1999). This remarkable finding was viewed with considerable skepticism when it first appeared in the literature, but current evidence regarding brain development in humans (Clancy & Finlay, this volume) confirms that the human brain is "up and running" and capable of learning before the third trimester. A large and comprehensive body of research by Kuhl, Werker, Jusczyk, and others (see especially Jusczyk, this volume) testifies to the rich and intricate patterns of language-specific speech contrasts (i.e., phonotactics) that infants develop during the first year of life, zeroing in like homing pigeons first on the vowels favored by their language, and then on consonantal boundaries, rhythmic biases, and constraints on the kinds of sounds that can and cannot occur together.

The paper by Gerkin and Gomez reviews exciting new evidence suggesting that human infants (by at least 6–8 months of age) are so skilled at learning that they are able to pick up statistical patterns in their perceptual input with 2 minutes or less of exposure to a disembodied voice, played while the infants are playing on the floor and seemingly paying little attention. They go on to show that the same statistical learning process that operates on speech (locating possible word boundaries) are also capable of supporting the induction of artificial grammars with many of the properties that underlie the natural grammars that infants will acquire many months later. Gerkin and Gomez also cite studies by other investigators showing that infants can extract these same kinds of patterns from other kinds of perceptual inputs as well, for example, sequences of arbitrary tones or even lights (e.g., Saffran et al., 1999). This suggests the existence of a very general perceptual pattern extractor – not one specific to language – from relatively early in infancy. Moreover, in more recent work Ramus et al. (2000) have shown that tamarin monkeys can extract patterns from speech in exactly the same way as human infants – suggesting in this case that the pattern extractor is not even specfic to humans. We thus see, once again, that human language evolved to fit with pre-existing primate (or mammalian) perceptual processes, not the other way around.

In this fashion, the center of gravity in research on infant speech perception has shifted from what children are able to perceive (which seems to be just about everything we can throw at them) to what they are (eventually) unable to perceive. As adults, native speakers of Japanese find it extremely difficult to perceive the contrast between \ra\ and \la\ – a contrast that is very easy for listeners in a language like English, in which that contrast is used productively to distinguish between possible words. In the same vein, native speakers of English lose the ability to hear a Thai consonant contrast that falls somewhere between the English consonants \d\ and \t\ (see Werker & Desjardins, this volume). Interestingly, this loss starts around 10 months of age – right around the age at which children start to show systematic evidence of word comprehension. In order to tune into one's native language and find those packages of sound that really matter, human infants have to learn to "tune out" the array of sound contrasts that matter far less in their particular language. But there is an interesting twist: the infant's open mind about the languages she can learn is closed not by some mysterious maturational process depending on an invariant and inflexible "critical period," but by the very act of learning her own native language. Tuning in involves tuning out. The gradual (but very efficient) process of tuning in to possible and actual words in the speech stream is illustrated very clearly in the paper by Fernald et al., who use infants' looking behavior as an index of the strength, speed, and efficiency of word recognition across the second year of life.

References

Aslin, R.N., Pisoni, D.B., & Jusczyk, P.W. (1983). Auditory development and speech perception in infancy. In P.H. Mussen (Series Ed.) & M.M. Haith & J.J. Campos (Vol. Eds.), *Handbook of child psychology: Vol. II. Infancy and developmental psychobiology* (4th edn, pp. 573–687). New York: John Wiley & Sons.

Cutting, J.C., & Rosner, B.S. (1974). Categories and boundaries in speech and music. *Perception & Psychophysics*, *16*, 564–570.

DeCasper, A.J., Lecanuet, J.P., Busnel, M.C., Granier-Deferre, C., & Maugeais, R. (1994). Fetal reaction to recurrent maternal speech. *Infant Behavior and Development, 17* (2), 159–164.

Eimas, P.D. (1975). Auditory and phonetic coding of the cues for speech: Discrimination of the (R-L) distinction by young infants. *Perception & Psychophysics, 18* (5), 341–347.

Kuhl, P.K., & Miller, J. (1975). Speech perception by the chinchilla: Voiced-voiceless distinction in alveolar plosive consonsants. *Science, 190*, 69–72.

Lotto, A.J., Kluender, K.R., & Holt, L.L. (1997). Perceptual compensation for coarticulation by Japanese quail (Coturnix coturnix japonica). *Journal of the Acoustical Society of America, 102* (2), Pt. 1: 1134–1140.

Ramus, F., Nespor, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition, 73* (3), 265–292.

Ramus, F., Hauser, M.D., Miller, C., & Morris, D. (2000). Language discrimination by human newborns and by cotton-top tamarin monkeys. *Science, 288* (5464), 349–351.

Saffran, E.M., Aslin, R.N., & Newport, E.L. (1996). Statistical learning by 8-month-old infants. *Science, 274*, 1926–1928.

Saffran, E.R., Johnson, E.K., Aslin, R.N., & Newport, E.L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition*, *70*, 27–52.

ONE

Finding and Remembering Words: Some Beginnings by English-Learning Infants

Peter W. Jusczyk

Learning to speak and understand a language involves acquiring a vocabulary. At a minimum, the learner must be able to recognize and identify a set of sound patterns and attach these to their appropriate meanings. If the sound patterns were presented one at a time, or with clear pauses between adjacent words, then the first step would be a matter of discriminating and remembering the different patterns. However, in everyday speech, talkers rarely pause between words. Instead, they are more apt to run one word into the next. Adults are rarely aware of this difficulty in their own language because they have learned to use cues to word boundaries that are available in the speech signal. The word segmentation problem typically becomes apparent when one listens to speech in an unfamiliar language. Then, it becomes difficult to know where one word ends and the next one begins. Because different languages cue word boundaries in different ways, learners must discover the cues that are most useful in segmenting words for their native language.

Although infant-directed speech is generally slower and has more exaggerated pitch contours than adult-directed speech (Fernald & Simon, 1984), the lack of clear boundaries between successive words is still present. Nevertheless, the word segmentation problem might not pose a serious difficulty for acquiring a vocabulary if new words were always presented to infants as isolated utterances. However, even when parents are explicitly instructed to teach their children new words, they present these words in isolation only about 20% of the time (Woodward & Aslin, 1990). Consequently, to make real headway in acquiring a vocabulary, learners need to solve the word segmentation problem.

Potential Cues to Word Boundaries

There are several different sources of information that could potentially inform listeners about likely word boundaries in fluent speech. One possibility is that listeners rely on their knowledge of predominant word stress patterns. This notion figures prominently in the Metrical Segmentation Strategy proposed by Cutler and her colleagues

(Cutler, 1994; Cutler & Carter, 1987; Cutler & Norris, 1988). Noting that a very high proportion of content words in English conversational speech are stressed on their initial syllables, Cutler and her colleagues have suggested that, as a first-pass strategy, listeners might identify the potential onsets of words with the occurrence of stressed syllables.

Phonotactic constraints (restrictions on the permissible sequences of phonetic segments in words) have also been suggested as a potential source of information about word boundaries (Brent & Cartwright, 1996; Myers et al., 1996). For example, knowledge that English does not allow certain sequences of consonants, such as "db" or "kt," at the beginnings of words could be used to infer a potential word boundary between such consonants.

In the linguistics literature, the fact that certain allophones (different pronunciations of the same speech sound) are restricted to particular contexts has been suggested as a possible cue to word boundaries in speech (Bolinger & Gerstman, 1957; Hockett, 1958). For example, as Church (1987) has noted, in English, "t's" are aspirated (i.e., produced with a large puff of air) when they occur in the initial position of stressed syllables, but unaspirated elsewhere. Thus, a listener sensitive to the occurrence of an aspirated "t" in speech might infer that it marks the beginning of a new word.

Finally, it has also been suggested that distributional evidence (i.e., information about the kinds of contexts that a particular sound pattern appears in on different occasions) can serve as a cue to a word boundary (Brent & Cartwright, 1996; Saffran et al., 1996; Suomi, 1993). For example, hearing the word "milk" in a variety of different contexts (e.g., "the milk," "an old milk bottle") may help that word to "pop out" as a unit in the speech stream. Subsequently, the learner might be able to use this knowledge to infer information about word boundaries of two unfamiliar words (i.e., "chocolate" and "carton") in the sequence "chocolate milk carton."

It should be noted that none of these potential cues is completely reliable by itself in predicting word boundaries in English. Rather, each of them points to probable locations of word boundaries, and it is likely that in segmenting words from speech, listeners rely on some appropriately weighted combination of these cues.

When Does Word Segmentation Begin?

Investigating the word segmentation abilities of infants requires the use of a test procedure that allows for the presentation of long strings of speech. Aslin and I adapted the head-turn preference procedure for this purpose (Jusczyk & Aslin, 1995). We familiarized English-learning infants with pairs of words like "feet" and "bike." There were 15 different versions of each word. On a given familiarization trial, one of the words began to play when the infant looked at a flashing red light on one of two side panels. Repetitions of the word continued playing until either the trial was completed (i.e., after all 15 versions were played) or the infant turned away from the light for two consecutive seconds. At the completion of a familiarization trial, a green center light began to flash to attract the infant's attention to the center. Then the next trial began with one of the two red side lights flashing. This familiarization procedure continued until the infant accumulated at least 30 s of listening time to each word. Then, the infants heard four different passages, each consisting of six sentences. Two of the passages contained one of the familiarized words in each sentence (e.g., a "feet" passage and a "bike" passage), and two similarly contained repetitions of two words not heard during familiarization. On a given trial, as during the familiarization period, the test passage began to play when the infant was looking at the flashing red light. The passage either continued to its conclusion or was stopped when the infant turned away from the light for two consecutive seconds. Estimates of listening times to each passage were based on how long the infants looked at the flashing light per trial.

When 7.5-month-olds were tested with this procedure, they listened significantly longer to the passages that contained the words that they had been familiarized with, suggesting that they detected the occurrence of the familiarized words in these passages. By comparison, 6-month-olds tested with the same materials displayed no significant preferences for the passages with the familiarized words. Hence, in English-learners, the ability to detect familiar words in fluent speech appears to develop between 6 and 7.5 months of age.

How Does Word Segmentation Begin in English-Learners?

Now that we know when word segmentation abilities develop, the next issue to resolve concerns the means by which infants accomplish this task. Previous work has shown that infants' sensitivity to predominant word stress patterns and to phonotactic constraints in the native language increases between 6 and 9 months of age (Echols et al., 1997; Friederici & Wessels, 1993; Jusczyk et al., 1993a; Jusczyk et al., 1993b; Morgan & Saffran, 1995). These findings suggest that such sources of information may be available to infants in segmenting words from speech.

My co-workers and I have recently focused our investigations on whether infants might rely on some form of the Metrical Segmentation Strategy (Cutler & Norris, 1988). Using the same procedure as Aslin and I did, we familiarized 7.5-month-olds with pairs of words that each had an accented first syllable followed by an unaccented second syllable (Houston et al., 1995; Newsome & Jusczyk, 1995). These *strong-weak* words included "doctor" and "candle" (or "kingdom" and "hamlet"). Following the familiarization phase, the infants were tested on passages either with or without the familiarized target words. The infants listened significantly longer to the passages containing the target words.

One interpretation of these results is that 7.5-month-old English-learners can segment words with strong-weak stress patterns from fluent speech. However, another possibility is that the infants were responding only to the strong syllables of these words (i.e., not to "candle," but to "can"). To explore this possibility, we ran another experiment in which infants were familiarized with just the isolated strong syllable of each word (i.e., "dock" and "can" or "king" and "ham") and then heard the passages containing the original strong-weak words. The infants did not listen longer to the passages with the strong-weak words (e.g., "hamlet" and "kingdom") that corresponded to the strong syllables from the familiarization period (e.g., "ham" and "king"). Nor did infants familiarized with isolated strong-weak words like "hamlet" or "kingdom" listen longer to fluent speech passages containing the words "ham" or "king."

Further experimentation indicated why infants recognized the whole word instead of just the embedded strong syllable. One consequence of using a strategy that identifies stressed syllables with the onsets of words in speech is that words beginning with unstressed syllables would be missegmented. To examine whether learners encounter such difficulties, we conducted comparable experiments using weak-strong words (i.e., an unaccented syllable followed by an accented one). Thus, infants were familiarized with “guitar” and “surprise” (or “beret” and “device”). In contrast to the earlier results, 7.5-month-olds familiarized with weak-strong words gave no evidence of subsequently recognizing these words in sentential contexts. However, infants familiarized with just the strong syllables of these words (i.e., “tar” and “prize”) did listen significantly longer to the passages containing the whole weak-strong words (i.e., “guitar” and “surprise”). It was as if the infants perceived the “tar” from “guitar” as initiating a new word when it occurred in a fluent speech context.

Why do infants match familiarized strong syllables to words they hear in the test passages in the case of weak-strong words, but not in the case of strong-weak words? The distributional properties of the sentential contexts apparently are the key. Whenever a strong-weak word occurred in a sentence, its strong syllable was always followed by the same weak syllable (i.e., the one in the word). This was not true for the strong syllable of a weak-strong word. For example, the “tar” of “guitar” was followed by “is” on one occasion, by “has” on another, and by a sentence boundary on another. These differences across the various sentential contexts may help to signal a word boundary at the end of “guitar.” Indeed, Saffran et al. (1996) found that 8-month-olds can use distributional cues to segment wordlike patterns from strings of nonsense syllables. In the present case, the strong-syllable segmentation strategy posits a word onset at the strong syllable “tar.” This, plus the distributional evidence, makes “tar” pop out of the context as a word. To test this hypothesis, we rewrote our sentential materials to use a constant word following a particular target word. For example, “guitar” was always followed by “is,” and “surprise” was always followed by “in.” This time, when 7.5-month-olds were familiarized with the isolated syllables “tar” and “prize,” they did not listen significantly longer to the passages with “guitar” and “surprise.” One suggestion is that the context led them to segment the pseudowords “taris” and “prizin.” This interpretation was verified when infants familiarized with “taris” and “prizin” did listen longer to passages that included the word sequences “guitar is” and “surprise in.” Thus, when the distributional context is favorable, using stressed syllables to mark word onsets may cause infants to missegment speech as containing a possible strong-weak word.

Our results suggest that English-learning 7.5-month-olds begin to segment speech by using the occurrence of strong syllables to indicate onsets of new words. Although this strategy is helpful for words beginning with strong syllables, it is problematic for words that begin with, or consist solely of, weak syllables. Further experiments that we have conducted with 10.5-month-olds suggest that by this age, infants have resolved their problems with weak-strong words by supplementing their initial strategy, using additional cues to word boundaries. For example, although 9-month-olds gave no evidence of using context-sensitive allophones in segmenting words, 10.5-month-olds are able to use these kinds of cues (Jusczyk et al., 1998). Picking up these additional sources of information may be facilitated by the use of stress-based cues to break the input into smaller sized chunks, which may provide the learner with more opportunities to detect the correspondence between certain allophones (and also phonotactic patterns) and their relation to the onsets and offsets of possible words in the speech stream. Thus, the learner may progress by a "divide and conquer" strategy of segmenting utterances into smaller pieces and then tracking regularities within these.

Remembering Words

Segmenting words from speech will be of little help in building a vocabulary unless learners encode and remember the sound patterns of these words. There is evidence that even at 7.5 months of age, infants are storing information about sound patterns that they hear frequently. For example, in one study (Jusczyk & Hohne, 1997), 8-month-olds were visited 10 times during a 2-week period. On each occasion, they heard audio recordings of the same three children's stories (although by different talkers or in different orders on different days). Two weeks after the last home visit, the infants were brought into the laboratory, and lists of words were played. Half of the lists contained words that had occurred frequently in the stories (these words were new examples of the words, spoken in isolation and recorded separately from the stories). The other lists were made up of foils – words that had not appeared in the stories. The foil words occurred with the same typical frequency in child-directed speech as the story words and had phonetic properties similar to those of the story words. The infants who had heard the stories listened significantly longer to the story words than to the foils. By comparison, a control group of infants who had not heard the stories showed no preference for either type of word list. Thus, the results suggest that the infants who had heard the stories did segment and remember, over a 2-week period, some of the frequently occurring words in the stories.

Because recognition of the story words was indexed by an overall preference for lists of words from the stories, we cannot say whether the infants remembered all the words or just a few of the words that occurred on the lists. In an effort to obtain information about infants' memory for specific items, Houston, Tager, and I adapted the procedure and test materials Aslin and I had used to study when infants develop word segmentation abilities (Houston et al., 1997). In our first experiment, 7.5-month-olds were familiarized for 30 s each to either "feet" and "bike" or "cup" and "dog." The next day, the infants were tested on four passages, each of which used one of these words in every sentence. Even after the 24-hr delay, the infants listened significantly longer to the passages that contained the familiarized words. Hence, infants do appear to remember these specific sound patterns for at least a day.

Results of additional studies suggest that at this age, long-term memory for words may be closely tied to characteristics of the talker's voice. When testing immediately followed familiarization, infants listened longer to passages containing the familiarized words even when the isolated words were produced by one talker and the passages were produced by a different talker of the same gender. But this generalization across talkers failed when a 24-hr delay intervened between familiarization and testing. Under these circumstances, the infants were just as likely to listen to the passages with the novel words as they were to listen to those with the familiarized words. These findings suggest that, at least initially, infants' representations of words may be stored exemplars of previously heard words rather than abstract prototypes.

Conclusions

The studies reviewed provide some indication that the lexicon begins to develop relatively early in the second half of the 1st year. English-learners display some capacity for segmenting words from fluent speech by about 7.5 months of age. These earliest attempts at word segmentation appear to draw on information about predominant word stress patterns and distributional cues. This first-pass strategy succeeds in correctly segmenting many words, but not others. Yet, the strategy may also facilitate the acquisition of information about other potential cues (e.g., context-sensitive allophones and phonotactic constraints) to word boundaries in the language. Even at this early stage of language development, there is evidence that infants retain information about words that occur frequently in the input. At the same time, there is some evidence that these early memory representations of words are relatively limited. Further experience may be required to generalize from words produced by one talker to those produced by a different talker.

Acknowledgments

The author is grateful to Cynthia Fisher and Karla Jusczyk for comments on a previous version of the present manuscript, and to his many collaborators on the research projects reported here.

References and Further Reading

Bolinger, D.L., & Gerstman, L.J. (1957). Disjuncture as a cue to constraints. *Word, 13*, 246–255.

Brent, M.R., & Cartwright, T.A. (1996). Distributional regularity and phonotactic constraints are useful for segmentation. *Cognition, 61*, 93–125.

Church, K.W. (1987). *Phonological parsing in speech recognition.* Dordrecht, The Netherlands: Kluwer Academic.

Cutler, A. (1994). Segmentation problems, rhythmic solutions. *Lingua, 92*, 81–104.

Cutler, A., & Carter, D.M. (1987). The predominance of strong initial syllables in the English vocabulary. *Computer Speech and Language, 2*, 133–142.

Cutler, A., & Norris, D.G. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance, 14*, 113–121.

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

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

Friederici, A.D., & Wessels, J.M.I. (1993). Phonotactic knowledge and its use in infant speech perception. *Perception & Psychophysics, 54*, 287–295.

Hockett, C.F. (1958). *A course in modern linguistics.* New York: Macmillan.

Houston, D., Jusczyk, P.W., & Newsome, M. (1995, November). *Infants' strategies of speech segmentation: Clues from weak/strong words.* Paper presented at the 20th Annual Boston University Conference on Language Development, Boston.

Houston, D., Jusczyk, P.W., & Tager, J. (1997, November). *Talker-specificity and the persistence of infants' word representations.* Paper presented at the 22nd Annual Boston University Conference on Language Development, Boston.

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

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

Jusczyk, P.W., Cutler, A., & Redanz, N. (1993a). Preference for the predominant stress patterns of English words. *Child Development, 64*, 675–687.

Jusczyk, P.W., Friederici, A.D., Wessels, J., Svenkerud, V.Y., & Jusczyk, A.M. (1993b). Infants' sensitivity to the sound patterns of native language words. *Journal of Memory and Language, 32*, 402–420.

Jusczyk, P.W., & Hohne, E.A. (1997). Infants' memory for spoken words. *Science, 277*, 1984–1986.

Jusczyk, P.W., Hohne, E.A., & Bauman, A. (1998). Infants' sensitivity to allophonic cues for word segmentation. *Perception & Psychophysics, 61*, 1465–1476.

Morgan, J.L., & Saffran, J.R. (1995). Emerging integration of sequential and suprasegmental information in preverbal speech segmentation. *Child Development, 66*, 911–936.

Myers, J., Jusczyk, P.W., Kemler Nelson, D.G., Charles Luce, J., Woodward, A., & Hirsh-Pasek, K. (1996). Infants' sensitivity to word boundaries in fluent speech. *Journal of Child Language, 23*, 1–30.

Newsome, M., & Jusczyk, P.W. (1995). Do infants use stress as a cue for segmenting fluent speech? In D. MacLaughlin & S. McEwen (Eds.), *Proceedings of the 19th Annual Boston University Conference on Language Development* (Vol. 2, pp. 415–426). Somerville, MA: Cascadilla Press.

Saffran, J.R., Aslin, R.N., & Newport, E.L. (1996). Statistical learning by 8-month-old infants. *Science, 274*, 1926–1928.

Suomi, K. (1993). An outline of a developmental model of adult phonological organization and behavior. *Journal of Phonetics, 21*, 29–60.

Woodward, J.Z., & Aslin, R.N. (1990, April). *Segmentation cues in maternal speech to infants.* Paper presented at the biennial meeting of the International Conference on Infant Studies, Montreal, Quebec.