When listening to speech in an unfamiliar language, many of us have experienced difficulty hearing where one word ends and another begins. We might attribute this to the fact that speakers of this language speak more rapidly than we do. However, many non-native speakers voice similar complaints when listening to English. As fluent speakers of a language, we seem to have little difficulty in finding the beginnings and endings of words spoken in our native tongue, so why should it be so difficult to do so for a foreign language?

The answer lies in how words in conversational speech are typically produced: words are run together without reliable pauses between them. What native listeners have learned to do is to use their knowledge of certain regularities in the sound structure of the language to predict the boundaries of words. Because the nature of these regularities is tied to the particular sound structure of the language, knowledge of such regularities in one language is not always helpful in predicting word boundaries in another language. Thus, for infants to become experts at segmenting words in their native language they need to learn the particular features of the sound structure that are most predictive of word boundaries.

When do infants begin to segment words from speech?

Infants begin to show sensitivity to the sound organization of their native language during the second half of their first year\(^1\). During this period, sensitivity to non-native speech contrasts tends to decline\(^2\) and phonetic categories begin to organize along the lines of those in the adult language\(^3\). At the same time, infants appear to begin learning about how sounds typically form patterns in words in the language. For example, every language has ‘phonotactic’ constraints; that is, particular restrictions on which sequences of sounds are permissible in the words of the language. Thus, English does not allow words to begin with consonant sequences such as /zw/ or /vl/, but Dutch does. At six months, infants listen equally long to words with permissible or impermissible sequences for their native language, but by nine months, Dutch and English learners favor words with the permissible sound sequences\(^4\). Similar preferences develop between six and nine months for the prosodic characteristics of native language words. English-learning nine-month-old infants listen longer to words with the predominant (strong/weak) stress pattern than to ones with the less common (weak/strong) stress pattern\(^5\). Because the sound properties that infants are developing sensitivity to are potential sources of information about word boundaries\(^6-9\), it is natural to assume that word segmentation abilities might begin to develop at this time.

To determine when English-learning infants might begin to segment words from fluent speech, Jusczyk and Aslin\(^15\) familiarized infants for 30 s to each of a pair of words, such as *feet* and *bike* or *cup* and *dog*. The purpose of the familiarization was to prime the infants to respond to target words that would subsequently appear in some fluent speech passages. Jusczyk and Aslin predicted that infants would listen longer to passages that contained the familiarized targets than to passages without the familiarized targets.
They found that 7.5-month-old, but not six-month-old, infants did listen longer to the test passages with the familiarized targets, suggesting that the older infants recognized the occurrence of the targets in the fluent speech contexts (Fig. 1). In an additional experiment, 7.5-month-old infants familiarized with two passages containing target words listened longer to subsequent repetitions of these targets in isolation than to repetitions of comparable words not in the familiarization passages. Thus, even when infants’ initial exposure to the targets occurred in fluent speech passages, they showed some ability to segment the words from these contexts. Finally, infants familiarized with an item (e.g. tap) phonetically similar to a target word in the passage (e.g. cup) did not listen longer to these passages than to control passages. This indicates that infants respond to a detailed representation of the target words rather than to just a salient acoustic feature, such as their vowel quality. Finally, a subsequent investigation suggests that these segmentation abilities contribute to the development of a lexicon. Houston et al. found that 7.5-month-old infants, familiarized with target words on one day and tested 24 h later, listened longer to passages with the targets than to ones without them. Hence, infants appear to encode information into memory about the sound patterns of words that occur frequently in speech directed to them.

How do infants segment words?

There are several different sources of information in the speech signal that could be helpful to infants in segmenting words (Box 1). A number of recent investigations have focused on the ability of English-learning infants between seven and 11 months of age to use one or more of these sources of information in word segmentation. Interest in the possibility that infants might use a stress-based strategy in segmenting words was prompted by the fact that they first display sensitivity to the predominant stress pattern of English words (strong/weak as in fall/fallen) at some point between six and nine months of age. To determine whether infants use such information in word segmentation, Jusczyk et al. examined how English learners segment words with and without the predominant stress pattern. They found that 7.5-month-old infants correctly segmented bisyllabic words with the predominant stress pattern (i.e. strong/weak), but not words with a less frequent stress pattern (i.e. weak/strong). Specifically, 7.5-month-old infants familiarized with words such as kingdom and candidate listened longer to passages containing these words than to control passages. By comparison, 7.5-month-old infants familiarized with words with weak/strong stress patterns, such as desire and guitar, did not give evidence of detecting these words in passages (Fig. 2). Instead, 7.5-month-old infants appeared to mis-segment the weak/strong words at the strong syllable boundary. Hence, when familiarized with tar and rice, they listened longer to passages containing guitar and desire than they did to control passages. The same general pattern of greater success in finding targets with strong/weak, as opposed to weak/strong, stress patterns in fluent speech contexts was also noted for nine-month-old infants’ abilities to detect a familiarized two-syllable pattern in a four-syllable context. However, by 10.5 months, English learners do detect familiarized weak/strong words in fluent speech contexts. This suggests that by 10.5 months, English learners do not rely exclusively on stress cues to segment words from fluent speech.

English-learning eight-month-old infants are also capable of exploiting statistical regularities in the input as word segmentation cues. Saffran et al. exposed eight-month-old infants to a two-minute string of continuous synthetic speech composed of four different three-syllable sequences produced with flat stress. The order of the syllables within a sequence was fixed (e.g. tibudo, pabiku), however, each such three-syllable word was followed equally often by one of the three other ‘words’. Thus, within a word like tibudo, the probability that tibu followed tibu was 1.0, which was similar to the likelihood of /tib/ following /tib/. However, across word boundaries, the probability of a particular syllable following the last syllable of the preceding word was only 0.33. The listeners presumably segmented the words from the familiarization sequence (e.g. inido and pisp), together with two ‘part-words’ composed of the last syllable of one word plus the first two syllables of another word from the familiarization sequence (e.g. inindo and pispal). Note that during the familiarization sequence, the probability of /in/ following /in/ and of /pis/ following /pis/ was only 0.33. The listeners presumably produced a word from the unfamiliar sequence indicated that the eight-month-old infants did distinguish the words from the part-words. In particular, they treated the part-words as novel items (Fig. 3). Hence, when such statistical regularities are present in the input, infants are able to use this information to segment possible words from a stream of speech.

There is also evidence that by nine months of age, English learners have begun to determine the way that phonotactic sequences line up with word boundaries in their language. For example, particular sequences of two consonants (i.e. CC sequences) might be more likely to occur in the same word than they are within words in English. Other CC sequences might be more common within words (e.g. /g/ than between words). Many et al. tested whether
nine-month-old infants are sensitive to the distribution of such sequences within and between words. In one experiment, they presented nine-month-old infants with two types of lists of CVCCVC items. For both types of lists, the CVCCVC sequences belonged to the internal CC sequence. For one type of list, the internal CC sequence was one that occurs in the list types concerned the internal CC sequence. For English words, the proponent pattern involves stress in the initial syllable of content words used in conversational speech (Ref. 6). On the basis of this finding, Cutler and Norris (Ref. 8) suggested that listeners might use a contextual segmentation strategy (MSS), whereby they identify word-syllable onsets with the occurrence of strong (imposed) syllables in fluent speech. Evidence from a number of investigations suggests that adult English listeners do in fact segment words this way (Refs 9, 10). However, infants familiarized with isolated repetitions of two words and tested on passages that either included (shaded columns) or did not include (open columns) these items. The left-hand panel shows results for words with strong/weak stress (Ref. 7). The right-hand panel shows results dealing with the main focus of the present investigation, the effects of stress pattern on syllable segmentation. The distributional cues are weak/strong. Thus, match two syllables /kp/ and /yy/ greater than the one between /yy/ and /kp/ because /kp/ can be followed by many other words (e.g. ‘happy man’, ‘happy dog’, etc.). Thus, matching known lexical units to the input could help in isolating other words from fluent speech.

References
m. Holmgren, B.L. and Gromen, L. (1997) Structure as a cue to constraints Word. 13, 246–255
o. Laffoon, J.R., Newbrooke, E.L. and Adle, R.H. (1986) Word segmentation: the role of...
words nitrate and ‘night rates’ to detect these words in fluent speech contexts. They familiarized infants with isolated versions of one of these words and another word (either doctor or bundle) and then tested them on passages that either included or did not include these targets. Although an earlier investigation had shown that two-month-old infants can discriminate the allophonic differences between nitrate and night rates, nine-month-old infants gave no indication of using this information to locate the familiarized target word in the passage. Hence, nine-month-old infants familiarized with nitrate listened equally long to the test passage with night rates as they did to the one with nitrate. In contrast, 10.5-month-old infants did listen significantly longer to the test passage that contained the familiarized item. Thus, sensitivity to allophonic cues are distributed within words seems to develop in English learners between nine and 10.5 months.

Why multiple cues are necessary for word segmentation

The studies reviewed above indicate that, towards the end of the first year, English learners are sensitive to a number of different possible sources of information about word boundaries in fluent speech. This is fortunate because none of these sources is sufficient for correctly segmenting all words from fluent speech. For example, a complete reliance on prosodic cues, as in metrical segmentation strategy (Box 1), would lead an English listener to mis-segment the names of words beginning with weak (unstressed) syllables. Similarly, reliance on statistical regularities without consideration of other speech cues could cause a listener who knows the word candle to make segmentation errors in contexts such as *can deliver* or *can deliver three boxes*. Likewise, although /nt/ occurs relatively infrequently within words, this sequence does occur in finance. Consequently, listeners must draw on some combination of these potential cues in segmenting words from English speech.

Although more empirical research is needed to confirm the developmental picture, it appears that stress-based and statistical cues are available earlier for English learners than are phono-

tactic and allophonic cues. One possible reason for this progression is that infants need to perform at least a rough partitioning of the input into word-sized chunks to learn how the phonotactic and allophonic cues are distributed with respect to word boundaries. In any case, as infants gain access to a larger set of possible word segmentation cues, the question arises as to how infants integrate these different sources of information. How do infants weigh these different sources? Are some cues treated as more reliable indicators of word boundaries than others, or are the various cues summed in some way?

Many recent models of word recognition have attributed an important role to existing items in the lexicon in recovering words from fluent speech (see Ref. 28 for a review; Refs 29–31). Thus, in the long run, many of the potential cues to word boundaries might be used primarily in ruling out alternative parses of the speech signal.

Segmenting words when extracting meanings

The studies reviewed thus far demonstrate that, towards the end of their first year, infants have the ability to detect the sound patterns of familiarized words embedded in fluent speech. However, the fact that 7.5-month-old infants might recognize the occurrence of kingdom in a passage does not entail that they attach any meaning to this sound pattern. Ultimately, to comprehend sentences, infants will have to recover the meanings of words that they segment from utterances. Although one might expect that infants practiced in extracting sound patterns will smoothly transfer this ability to situations in which they must respond to the meanings of words, this does not appear to be the case.

Fernald et al. tested English-learning 15-month-old infants in a task in which a target word was embedded in a sentence. Specifically, infants were presented with two objects displayed on video monitors while they heard a sentence including the name of one of the objects. Although 15-month-old infants looked significantly more often at the picture of the named target when it occurred in the final position of the sentence, they did not do so when the target occurred in the middle of the sentence. The finding that the positioning of the target word in the sentence matters for 15-month-old infants is important because studies with 7.5-month-olds have typically used the sentential position of the target in test passages (29–31) and found no response bias (although Aslin has found some evidence for an utterance final bias in eight-month-olds; R.N. Aslin, unpublished data). In any case, Fernald et al. found that 18-month-old infants responded equally well to the targets in the medial and final positions of sentences. Why, then, do 15-month-old infants have difficulty with targets in non-final positions? One possibility is that the additional processing demands of associating a word to the correct picture might tax the word-segmentation abilities of 15-month-old infants. Fernald et al. suggest that placing the word in the final position might increase its salience for the infants, thereby allowing them to segment it. Some indirect support for this view comes from an investigation of word learning in 14-month-old infants by Stager and Werker. They found that the increased demands associated with a word-learning task had a detrimental effect on infants’ speech discrimination capabilities. Indeed, such declines in perceptual performance might occur whenever
infants move from a task requiring only sound processing to one that requires infants to use the extracted sound features to guide some additional response.

Infants might need time and practice to coordinate the routines used for segmenting speech with those required for associating sound patterns with their correct meanings. It has been reported that, by 18 months, English learners do not show any significant decline in perceptual sensitivity to this task (D.J. Swingley and R. N. Aslin, unpublished data). Other research by Fernald and her colleagues suggests that, as infants develop, they become faster and more accurate in responding to words in a setting where spoken words are associated with pictures. By carefully analyzing infants’ reaction times to the correct picture after hearing the target word, Fernald et al. documented that 24-month-old infants were 316 ms faster than 15-month-olds and 148 ms faster than 18-month-olds in shifting their gaze from a distractor picture to the target picture (Fig. 4). In other words, the older group was much faster at understanding the words they heard.

The first 24-month-old infants achieve more adult-like performance levels in recognizing words in these contexts was recently confirmed in a follow-up study by Swingley et al. By systematically manipulating the phonetic similarity of distractor items, they found that 24-month-old infants’ latencies to fixate the labeled picture were delayed when the distractors and targets overlapped phonetically at onsets, but not when they only overlapped at offsets. Adults tested in the same protocol showed a very similar pattern of results. Thus, it is clear that infants’ abilities to recognize particular words in fluent speech contexts undergo considerable improvement between 15 and 24 months. Conclusions

English-learning infants first demonstrate some capacity for word segmentation at about 7.5 months of age. Initially, they appear to rely on prosodic and statistical information to locate words in fluent speech. Although these sources of information are useful for segmenting content words with the predominant English stress pattern, they will also lead to missegmentations of words with other stress patterns. Nevertheless, this early word segmentation strategy might help infants to learn about the way that other potential word boundary cues relate to the phonostructural and allophonic properties of the language. Ultimately, then, language learners must draw upon multiple cues to determine word boundaries in fluent speech. Although infants’ word segmentation skills improve during the first year, there is evidence that they undergo further improvements during the second year. The task of having to attach meanings to sound patterns affects infants’ abilities to segment words. However, there is evidence that, as infants approach the end of their second year, their recognition of familiar words in fluent speech begins to approach that of adults.

Acknowledgements

The preparation of this manuscript was facilitated by a research grant from NIH (5T32 HD007160) and by a senior scientist award from NICHD (K02 HD08467) to PWJ. The author wishes to thank Ann Marie Jusczyk, Richard Aslin and the anonymous reviewers for helpful comments made on a previous version of this manuscript.

References


Outstanding questions

• English learners appear to use the predominant stress pattern of English words as an initial segmentation strategy. Is this simply a general perceptual bias to prefer strong/weak sequences, or do words with this pattern appear frequently in isolated words spoken to English learners?
• How do language learners respond to conflicting cues about word boundaries in fluent speech? Do they treat some sources of information as more reliable indicators of word boundaries than others?
• When do infants show signs of integrating different sources of information about word boundaries in fluent speech?
• How is such information integrated?
• Do language learners of other languages show a similar course of development of word segmentation skills as do English learners? How are developing word segmentation strategies affected by the sound organization of a particular language? If stress-based cues are not useful for a given language, where do learners begin?
• When learners begin to process meanings, as well as sound patterns, how are word segmentation processes affected? Do learners repeat the developmental sequence manifested in their first attempts at word segmentation? For example, when English-learning infants begin to process meanings, is it easier for them to segment words with strong/weak stress patterns than ones with weak/strong stress patterns?
Jusczyk – Word segmentation in infants

11 Carrera, P. et al. (1997) Bootstrapping word boundaries: a bottom-up corpus-based approach to speech segmentation Cognit. Psychol. 31, 111–132
12 Church, K. (1987) Phonological parsing and lexical retrieval Cognit. Psychol. 18, 53–69
32 Norris, D.G. et al. (1993) The possible word constraint in the segmentation of continuous speech Cognit. Psychol. 28, 141–210
35 Ferrand, A. et al. (1998) Rapid gains in the speed of verbal processing by Infants in the second year Psychol. Sci. 9, 208–211
36 Sc Sparse, D., Pittis, J. and Ferrand, A. Continuous processing in word recognition at 24 months Cognitive Science (in press)

Now available online… Trends in Cognitive Sciences

The online version of Trends in Cognitive Sciences contains a simple search and browse facility that will enable you to view full-text versions of all articles published in the journal since January 1998. Browsing and searching the articles is straightforward and designed around the need for simple navigation. Articles can be viewed in three display formats: text-only, text with pop-up figures, or as PDF files for high-quality printing.

Personal print subscribers benefit from online access to Trends in Cognitive Sciences online by accessing the site using a unique subscription key. For information about how to obtain your key please visit http://www.trends.com

The subscription key is important and should be kept safe as it is your unique access code to your online journal for the remainder of the year. Institutional access is currently available through Elsevier’s ScienceDirect (http://www.sciencedirect.com) services.

Subscribers and non-subscribers alike will continue to have free access to full content lists, abstracts, search facilities and the e-mail contents alerting service. It is also possible to download full-text articles on a pay-per-view basis.

The following articles are just some of those available in the fully searchable archive:


I hope that you enjoy what Trends in Cognitive Sciences online can offer you now and look forward to offering you further benefits in the future.

Peter Collins
Editor