Lingua 92 (1994) 81-104. North-Holland

Segmentation problems, rhythmic solutions*

Anne Cutler**

Medical Research Council, Applied Psychology Unit, 15 Chaucer Road, Cambridge CB2 2EF, UK

The lexicon contains discrete entries, which must be located in speech input in order for speech to be understood; but the continuity of speech signals means that lexical access from spoken input involves a segmentation problem for listeners. The speech environment of prelinguistic infants may not provide special information to assist the infant listeners in solving this problem. Mature language users in possession of a lexicon might be thought to be able to avoid explicit segmentation of speech by relying on information from successful lexical access; however, evidence from adult perceptual studies indicates that listeners do use explicit segmentation procedures. These procedures differ across languages and seem to exploit language-specific rhythmic structure. Efficient as these procedures are, they may not have been developed in response to statistical properties of the input, because bilinguals, equally competent in two languages, apparently only possess one rhythmic segmentation procedure. The origin of rhythmic segmentation may therefore lie in the infant's exploitation of rhythm to solve the segmentation problem and gain a first toehold on lexical acquisition. Recent evidence from speech production and perception studies with prelinguistic infants supports the claim that infants are sensitive to rhythmic structure and its relationship to lexical segmentation.

1. Introduction: Using the lexicon in listening to speech

The lexicon, which is the focus of the present collection of papers, is assumed to be an essential component of every language user's linguistic apparatus. The contents of a lexicon are so patently language-specific that it goes without saying that a lexicon cannot be inborn – it must be acquired, on the basis of linguistic experience. Such experience usually consists of hearing utterances, which typically are many words in length. But we can safely

* Financial support is acknowledged from the Human Frontier Science Program. Thanks as always to the colleagues with whom the many and varied experiments summarised here were carried out: Sally Butterfield, David Carter, Giyoo Hatano, Peter Jusczyk, James McQueen, Jacques Mehler, Dennis Norris, Takashi Otake and Juan Segui. For discussions and other assistance with this paper thanks to Melissa Bowerman, Sally Butterfield, Mike Kelly, James McQueen, Jacques Mehler, Dennis Norris and Ann Peters.

** Anne Cutler is now at Max Planck Institute for Psycholinguistics, P.O. Box 310, 6500 AH Nijmegen, The Netherlands.

SSDI 0024-3841(93)E0039-A

assume that complete utterances are not what lexicons consist of. The number of utterances is potentially infinite, so that to store all the utterances we might ever hear our lexicon would also have to have infinite capacity. Even if we set an arbitrary length limit, the number of possible utterances is enormously large; for instance, Miller (1967) calculated that there must be at least 10^{20} possible English sentences of twenty words or less – a total which, he drily added, would take considerably longer than the estimated age of the earth to speak.

Instead, we assume that the contents of a lexicon consist of sound-to-meaning mappings in discrete chunks. (We can refer to lexical entries by the shorthand term 'words', although of course not all lexical entries necessarily correspond to what would be written as a single separate word. Some sub-word forms such as affixes or stem morphemes may well have lexical representation, as may particles which are conjoined with other words in writing; likewise, multi-word idiomatic expressions and frequently occurring phrases may be represented by a single entry.) Thus using a lexicon requires the separation of utterances into the lexically relevant chunks of which they are made up – producing speech requires the language user to string together lexical entries to make a whole utterance, and recognising speech requires division of an utterance into units which can be looked up in the lexicon. Likewise, acquiring a lexicon eventually involves acquiring the ability to use it in these ways.

The present contribution focusses on the very start of lexicon-building: how the infant might find out what words in the input language are like, and might assemble an initial stock of known words. The initial task is perceptual. What exactly does it involve? For instance, does it involve (as mature use of a lexicon in speech recognition involves) division of multi-word utterances into lexically relevant chunks? And if so, how difficult is this task? To answer these questions we need to consider the nature of the speech input with which the infant is most likely to be confronted. Comparative studies of various types of speech are considered in the next section.

2. Styles of adult-directed speech

Most of the speech any listener hears is spoken spontaneously – the speech signals which occur in the majority of everyday situations have been conceived and composed by their speakers even as they are uttered. For most listeners, spontaneous speech is encountered far more often than other styles such as rehearsed speech (heard in the theatre or on radio or television, for example), read speech (in news broadcasts, or, too often, in lectures) or computer-synthesised speech.

Whatever the style of speech, words in isolation occur only rarely – nearly all utterances are multi-word. A lot is known about the phonetics of multi-word utterances, and a fair summary of our knowledge is that words are strongly affected by the contexts in which they occur; moreover, these contextual assimilation processes operate to obscure word boundaries, with the result that there are few reliable cues in a continuous speech signal to where one word ends and the next begins. Klatt (1989) provides a telling overview of the problems which this causes for the lexical access process so essential for speech recognition.

Nevertheless, the majority of such phonetic studies have been conducted on speech produced in laboratory situations, which is normally read speech. Is this a fair representation of the speech which most listeners usually hear? Motivated by this question, speech scientists have undertaken a number of studies aimed at describing spontaneous speech, and because of the underlying motivation, most of the studies have been comparative: spontaneous speech has been contrasted with read speech. These studies have revealed systematic differences between the two types of speech. Some of these differences might render the listener's problems even worse in spontaneous than in read speech. For example, casual spontaneous speech is particularly prone to phonological elisions and assimilations (G. Brown 1977, Labov 1972, Milroy 1980) and to syntactic simplifications and, occasionally, incompleteness (Cheshire 1982, Labov 1972). Other differences, however, might make life easier for listeners to spontaneous speech. These are principally differences in the prosodic domain. Thus spontaneous speech tends to be produced at a slower rate than read speech (Barik 1977, Johns-Lewis 1986, Levin et al. 1982), and to have longer and more frequent pauses and hesitations (Barik 1977, Crystal and Davy 1969, Kowal et al. 1975, Levin et al. 1982) and shorter prosodic units (Crystal and Davy 1969).

Listeners can distinguish spontaneous utterances either from read speech (Levin et al. 1982, Remez et al. 1985, Blaauw 1991) or from rehearsed speech (Johns-Lewis 1987); their judgements are most likely based on prosodic aspects of the speech, because accuracy is still high when the speech extracts have been low-pass filtered (Levin et al. 1982), while the distinction can not be as accurately made on written versions of the text (Johns-Lewis 1987). Fluent spontaneous speech can be identified as accurately as disfluent (Blaauw 1991).

The prosodic differences between spontaneous and read speech have consequences for the way speech in each mode is processed by listeners. McAllister (1991) examined word recognition in spontaneous and read speech using the gating task, in which listeners hear successively larger fragments of a word. She found that word identification (in context) occurred earlier for a word stressed on the first rather than on the second syllable in spontaneous speech, but not in read speech. In a word-by-word gating study of spontaneous speech Bard et al. (1988) and Shillcock et al. (1988) similarly found that words containing strong syllables were easier to identify than words which were realised as weak syllables.

Mehta and Cutler (1988) investigated phoneme detection reaction time in spontaneous and read speech, and compared in particular the relative strength in the two speech modes of a number of previously established effects. They found no overall difference in response time between the two speech modes, and also no difference between the two modes on the one semantic variable in the study, the effects of the transitional probability of the target-bearing word. However, four other effects differed across modes. In read speech but not in spontaneous speech, late targets were detected more rapidly than early targets, and targets preceded by long words were detected more rapidly than targets preceded by short words. In contrast, in spontaneous speech but not in read speech, targets were detected more rapidly in accented than in unaccented words and in strong than in weak syllables.

Mehta and Cutler explained these differences in terms of prosodic differences between the two speech modes. The greater frequency of hesitations in spontaneous speech, for example, results in shorter prosodic units, which in turn reduces the average span over which rhythmic predictability will hold. So because prosodic units are long – generally clause-length – in read speech, but usually short in spontaneous speech, the opportunity for rhythmic prediction in the latter case is much smaller. Mehta and Cutler thus argued that position in the sentence is not, strictly speaking, what affects target detection time; rather, the effective variable is position in the prosodic unit. Similarly, because hesitations tend to be more frequent and longer in spontaneous speech, it is much more likely that a particular target-bearing word will be preceded by a hesitation in the spontaneous than in the read mode. Where a target is immediately preceded by a hesitation, any effects of incomplete processing of the previous word will be nullified by the extra processing time provided by the hesitation, so that effects of preceding word length, which are held to reflect just such processing hangovers from the preceding word, will be less likely. Finally, because accent patterns in spontaneous utterances were more varied and less likely to express default accenting than those in read utterances, and the acoustic differences between strong and weak syllables were greater in spontaneous than in read speech, there was greater opportunity for processing effects of both sentence accent and syllable stress to appear in the spontaneous than in the read speech; this would account for the finding of significant facilitation due to sentence accent and syllable stress in the former but not in the latter. The results of the gating studies described above provide similar evidence of the perceptual importance of syllable stress in spontaneous speech.

These findings speak to the majority case for speech processing. Most speech that adult listeners hear is spontaneously produced. Such speech is characterised by a fairly slow overall rate of speech, short prosodic units, frequent pauses and, in English, a clear opposition between strong and weak syllables. These factors affect the way the speech is processed.

3. The infant's speech environment

Is the infant's speech environment the same as that of an adult listener? In one respect it is not, because speech directed to the infant as chosen listener exhibits a pattern which is systematic enough to warrant calling it a separate speech style. This conclusion emerges from a number of studies which have investigated the characteristics of speech addressed to young children at various stages of development; again, the studies have mainly been comparative, with infant-directed spontaneous speech being typically contrasted with speech from the same speakers to adults.

In European languages, infant-directed speech tends to be spoken at a slower rate, to have more frequent stresses, shorter continuous sequences and longer pauses, and to be higher in pitch than adult-directed speech (Fernald and Kuhl 1987, Fernald and Simon 1984, Fernald et al. 1989, Garnica 1977, Stern et al. 1983). Slower rate, more frequent prosodic demarcation, and longer pauses, it will immediately be noted, are the primary factors which the review in the previous section revealed as distinguishing spontaneous from non-spontaneous speech. To the present author's knowledge these two literatures have not been directly compared; there would seem to be a defensible case, however, for considering adult-directed and infant-directed spontaneous speech in terms of a single continuum, with infant-directed speech occupying a more extreme position on most measures.

The one exception is that infant-directed speech is reported to have higher pitch and a wider fundamental frequency range (Fernald and Simon 1984, Garnica 1977). In contrast, the fundamental frequency range of spontaneous speech has been reported in at least some studies to be relatively narrow, at least in intimate conversation (Johns-Lewis 1986, Blaauw 1991). Pitch is a particularly important dimension of infant-directed speech, since the fact that infants prefer to listen to this style of speech (Fernald 1985) has been found to be principally due to its pitch characteristics (Fernald and Kuhl 1987, Sullivan and Horowitz 1983).

Ohala (1983, 1984) has argued that raised pitch is an ethologically universal signal of smallness, ingratiation and non-threatening attitude. From such a perspective it would be possible to argue that raised pitch might not be a phonologically relevant manipulation in speech to infants, but might simply arise from universal expression of affection or nurturance on the part of an adult to an infant. Against this conclusion, on the other hand, might be cited the more recent findings that the pitch manipulations found in infant-directed speech in American English and related languages are apparently *not* universal. Although rising contours predominate in infant-directed speech in the stress languages English (Sullivan and Horowitz 1983) and German (Fernald and Simon 1984), falling contours are more prevalent in the tone languages Mandarin (Grieser and Kuhl 1988) and Thai (Tuaycharoen 1978). In a comprehensive review of the literature on pitch in infant-directed speech, Shute (1987) concluded that pitch modifications are not only clearly not universal across languages, but may also differ within one language as a function of sex of the speaker, age of the child addressee, frequency of the speaker's interaction with children and other factors.

In fact a recognisable style of infant-directed speech is itself not universal, contrary to the confident expectations of researchers in the 70s that it would prove not only to be universal (Ferguson 1977) but absolutely necessary for successful acquisition (R. Brown 1977). It is now clear that there are cultures where infants are exposed to much normal adult speech but no speech in any special infant-directed mode (Heath 1983, Schieffelin 1985, Schieffelin and Ochs 1983). Even where infant-directed speech appears to conform to the pattern observed in English and like languages, this may not constitute a specialised mode; thus infant-directed speech in Quiche Mayan has relatively high pitch, but so, in this language, does adult-directed speech from the same informants (Bernstein-Ratner and Pye 1984).

Thus it is reasonable to conclude that infants in the earliest stages of language acquisition receive at least some of their input – and for some infants, perhaps all of their input – in a form that at least closely resembles normal spontaneous speech between adults. One of the characteristics of spontaneous speech, it will be recalled, is the high frequency of phonological elisions and assimilations (G. Brown 1977, Labov 1972, Milroy 1980). Some studies have reported that child-directed speech, too, is replete with such distorting processes (Bard and Anderson 1983, 1991; Shockey and Bond 1980), which is consistent with the view that this style of speech lies on a general continuum with adult-directed casual speech. Other studies, however, have reported *lower* frequency of distorting phonological transformations in speech to infants than in speech to adults (Bernstein-Ratner 1984a,b). In an attempt to resolve this apparent contradiction, Stoel-Gammon (1984) transcribed five hours of speech to one-year-olds; her results strongly support the view of a continuum, since she effectively discovered a continuum within her own data, from very clear articulation (e.g. release of word-final stop bursts, clear articulation of unstressed vowels) to very casual forms (frequent vowel reduction, omissions of whole syllables such as [sko] as a pronunciation of *let's go*). Stoel-Gammon concluded that the phonological characteristics of speech to children depend on such factors as contextual redundancy, the function of the individual utterance, and the situational context – the same factors that determine the phonological forms of adult spontaneous speech (Lieberman 1963, Cheshire 1982).

There is to my knowledge no evidence, from any culture, of a greater incidence of isolated words in speech to children than in other forms of speech. Even though phrases may be short, they are still phrases. Thus a speech segmentation problem, as described in the introductory section to this paper, seems to exist for the infant as for the adult language user. The speech that the infant hears is continuous; much of the speech of the infant's environment will be speech among mature language users; in perhaps a majority of cultures speech addressed specifically to the infant would form only a small proportion of the input; even then, such speech may not necessarily be clearly articulated. The problem is compounded for the infant by the necessity of compiling a lexicon, and this added difficulty does not trade off against reduced segmentation difficulty in the input. In fact, the scale of the segmentation problem in the structure of the input is remarkably similar for the infant and for the adult.

4. The segmentation problem for adult listeners

The adult listener typically hears continuous multi-word utterances, and must therefore segment the speech stream in order to understand them. The importance of this segmentation problem has long been acknowledged, and many, widely varying, experimental approaches have tackled the question (e.g. Hayes and Clark 1970, Wakefield et al. 1974, Pilon 1981). But the adult's situation differs from that of the child because as a mature language user the adult is already in possession of a lexicon. This has led some researchers to claim that the adult listener has no need of explicit segmentation procedures, since the successful recognition of a word will ensure that whatever immediately follows that word will be known to be word-initial. For example, Cole and Jakimik (1978) proposed that recognition of spoken utterances proceeds in strictly temporal order, and 'one word's recognition automatically directs segmentation of the immediately following word' (1978: 93). We could call such a model 'segmentation by default'.

On closer inspection, though, it becomes obvious that segmentation by default could not work, at least for English. Firstly, the model relies on listeners being indeed able to determine where a word ends. It is true that sometimes phonetic sequence constraints will be of use in this. Some sequences of phonemes – [au k], [m g], [ai S], for example – never occur word-internally in English (Lamel and Zue 1984, Harrington et al. 1987); thus a sentence like *How come Guy shaved?* should prove very easy to segment. But unfortunately such helpful sequences are rare. As McQueen and Cutler (1992) have recently demonstrated, the English vocabulary contains few totally distinct words; most long words have other words embedded in them (as *reconciliation* contains *wreck, reckon, sill, silly, ill,* etc.), while most short words occur within longer words (as *late* can be found in *latency, collate, belated, translatability,* etc.). Moreover, McQueen and Cutler's computations showed that the majority of such overlaps occur at the beginnings of the longer words, a particular problem for segmentation by default.

Of course, in typical English speech the majority of words are monosyllabic (Cutler and Carter 1987), which will certainly reduce the problems caused by such embeddings. But most monosyllabic English words do not become unique until at or after their final phoneme (Luce 1986); and in fact many words – especially monosyllabic words – can not be recognised until after their acoustic offset. Post-offset recognition has been demonstrated both with laboratory-produced (i.e. carefully read) speech (Grosjean 1985), and with spontaneously produced speech (Bard et al. 1988, Shillcock et al. 1988). If words cannot be recognised till *after* their ends then segmentation by default would lose its very basis.

Secondly, models such as segmentation by default are far from robust; they assume that prelexical processing of the speech signal will be accurate. But in practice speech signals are not always fully clear. Background noise, distance between speaker and listener, distortion of the speaker's vocal tract, foreign accents, slips of the tongue – all these, and similar factors, conspire to make the listener's phonetic interpretation task harder. A much more robust model is needed to account for what is obviously true, namely that human speech recognition is extremely successful even under noisy conditions or with previously unfamiliar voices or accents.

5. A solution for English: Rhythmic segmentation

In fact there is a good deal of evidence from human speech recognition in English that explicit segmentation procedures are employed by adult listeners. Cutler and Norris (1988) suggested that the characteristic rhythmic structure of English could form the basis for an effective segmentation procedure, because English speech shows a systematic relationship between rhythmic patterns and word boundary locations. The rhythm of English is based on stress, with syllables of the language being either strong or weak; strong syllables contain full vowels, while weak syllables contain reduced vowels (usually schwa). Cutler and Carter (1987) demonstrated that English lexical words are far more likely than not to begin with strong syllables – in a 33,000-word phonetically transcribed dictionary (the MRC Psycholinguistic Database: Coltheart 1981, Wilson 1988), 73% of all entries had strong initial syllables. But the frequency of occurrence of individual words differs widely; lexical, or content words, are sometimes very common (e.g. *people*), but more often are very rare (e.g. *peon, steeple*), while some words which in running speech are usually realised as weak syllables – grammatical, or function words, such as *of* or *the* – occur very frequently. Cutler and Carter examined a 190,000-word natural speech sample, the *Corpus of English Conversation* (Svartvik and Quirk 1980), using the frequency count of this corpus prepared by Brown (1984); they found that in this corpus 90% of the lexical words have strong initial syllables. However, the grammatical words in the corpus were actually in the majority, and they were virtually all weak monosyllables. Cutler and Carter computed that about three-quarters of all strong syllables in the sample were the sole or initial syllables of lexical words, while more than two-thirds of all weak syllables were the sole or initial syllables of grammatical words.

This means that a listener encountering a strong syllable in spontaneous English conversation would seem to have about a three to one chance of finding that strong syllable to be the onset of a new lexical word. A weak syllable, on the other hand, would be most likely to be a grammatical word. English speech should therefore lend itself to a segmentation procedure whereby strong syllables are assumed to be the onsets of lexical words. Cutler and Norris interpreted results of an experiment they ran as evidence for such a procedure. They used a task which they called word-spotting, in which listeners were asked to detect real words embedded in nonsense bisyllables; detection times were slower to the embedded word in, say, *mintayf* (in which the second vowel is strong) than in *mintef* (in which the second vowel is schwa). Cutler and Norris interpreted this as evidence that listeners were segmenting *mintayf* prior to the second syllable, so that detection of *mint* therefore required combining speech material from parts of the signal which had been segmented from one another. No such difficulty would arise for the detection of *mint* in *mintef*, since the weak second syllable would not be divided from the preceding material.

Further evidence for such a procedure was produced by Cutler and Butterfield (1992), who investigated the way in which word boundaries tend to be misperceived. In both spontaneous and experimentally elicited misperceptions they found that erroneous insertions of a word boundary before a strong syllable (e.g. *achieve* being heard as *a cheap*) and deletions of a word boundary before a weak syllable (e.g. *bird in* being heard as *burgling*) were far more common than erroneous insertions of a boundary before a weak syllable (e.g. *effective* being heard as *effect of*) or deletions of a boundary before a strong syllable (e.g. *were waiting* being heard as *awaken*). This is exactly what would be expected if listeners deal with the segmentation problem by assuming that strong syllables are likely to be word-initial, but weak syllables are not.

As Cutler and Norris point out, the strong syllable is defined by the quality of its vowel (full, in comparison to the reduced vowels of weak syllables); thus spotting strong syllables cannot provide a complete solution to the segmentation problem since word boundaries actually occur prior to the onset of syllables. A strong syllable spotter must be supplemented by some means of estimating actual syllable onset; Cutler and Norris suggest that more than one alternative realisation of such a device would be feasible. Assuming that a rhythmically based segmentation procedure is indeed practical, its advantages are considerable. For instance, such a procedure is obviously not going to be affected by the frequency of words embedded within other words in speech, or by the relative frequency of monosyllables versus polysyllables. Only where polysyllabic words contain strong syllables in non-initial position will the procedure produce a non-optimal result (i.e. it will signal a word boundary but this will be a false alarm). However, polysyllabic words with non-initial strong syllables occur relatively rarely (Cutler and Carter 1987), and in only a small minority of them will a false alarm actually produce a real word unrelated to the embedding word (e.g. *late* in *collate*; Cutler and McQueen, in press). Thus rhythmic segmentation is a relatively efficient procedure for English.

It is also quite robust – in fact, it is precisely with uncertain input that rhythmic segmentation proves particularly useful. Researchers in automatic speech recognition (e.g. Shipman and Zue 1982) have developed systematic representations of phonetic uncertainty, namely transcriptions in which only general classes of phoneme are provided (e.g. glide, nasal, stop consonant, etc.). Two studies using uncertain input of this kind have produced further evidence in favour of rhythmic segmentation. In the first study, Briscoe (1989) implemented four segmentation algorithms and tested their performance on a (phonetically transcribed) continuous input, using a 33,000-word lexicon. The algorithms postulated potential lexical boundaries: (a) at the end of each successfully identified word ('segmentation by default'); (b) at each phoneme boundary; (c) at each syllable onset; and (d) at each strong syllable onset (the rhythmic segmentation proposal). The measure of performance was the number of potential lexical hypotheses generated (the fewer the better). With completely specified phonetic input all algorithms naturally performed quite well. However, significant differences between the algorithms emerged when some or all of the input was phonetically uncertain; most affected were 'segmentation by default' and the phonemic algorithm, both of which generated huge numbers of potential parses of incomplete input. Far better results were produced by the algorithms which constrained possible word onset positions in some way, and the more specific the constraints, the better the performance: the rhythmic segmentation algorithm performed best of all with the uncertain input. In the second study, Harrington et al. (1989) compared the rhythmic segmentation algorithm with a segmentation algorithm based on permissible phoneme sequences (Lamel and Zue 1984, Harrington et al. 1987), using as a metric the proportion of word boundaries correctly identified in a 145-utterance corpus. With phonetically uncertain input, sequence constraints proved virtually useless, but the rhythmic segmentation algorithm still performed effectively (in fact it correctly detected more word boundaries in uncertain input than the phoneme sequence constraints had detected in completely specified input).

The efficiency and robustness of rhythmic segmentation therefore suggest that listeners profit from employing an explicit segmentation procedure of this kind. A striking fact about this procedure, however, is its language-specificity: as described for English, the procedure is based on stress rhythm, i.e. the opposition of strong and weak syllables. Clearly, it therefore cannot be a universal strategy, because many (indeed most) languages of the world do not have stress rhythm. However, all languages have rhythm – speech rhythm need not be stress-based. In the next section alternative forms of rhythmic segmentation are described, supported by the results from experiments in languages which do not have stress rhythm.

6. Rhythmic segmentation in French and Japanese

Mehler (e.g. 1981) and his colleagues (e.g. Segui 1984) have used a variety of psycholinguistic tasks to demonstrate processing advantages for syllables in speech comprehension. In one experiment, which launched a series of cross-linguistic comparisons, Mehler et al. (1981) had French subjects listen to lists of unrelated words and press a response key as fast as possible when they heard a specified word-initial sequence of sounds. This target was either a consonant-vowel (CV) sequence such as *ba-* or a consonant-vowel-consonant (CVC) sequence such as *bal-*. The words which began with the specified sound sequence had one of two syllabic structures: the initial syllable was either open (CV), as in *balance*, or closed (CVC), as in *balcon*. Mehler et al. found that response time was significantly faster when the target sequence corresponded exactly to the initial syllable of the target-bearing word than when the target sequence constituted more or less than the initial syllable. Thus responses to *ba-* were faster in *balance* than in *balcon*, whereas responses to *bal-* were faster in *balcon* than in *balance*. Mehler et al. claimed that this result supported a syllabically based segmentation strategy in speech recognition in French. Similarly, Segui et al. (1981) found that listeners are faster to detect syllable targets than to detect targets corresponding to the individual phonemes which make up those same syllables. Further evidence from French that polysyllabic words, whether they are heard in isolation or in connected speech, are analysed syllable by syllable came from studies by Segui (1984) and by Dupoux and Mehler (1990).

If speech segmentation in French proceeds syllable by syllable, there is an interesting parallel to the results from English reported in the previous section. Just as use of the opposition between strong and weak syllables in segmenting English exploits the English language's characteristic stress-based rhythmic pattern, so does use of the syllable in segmenting French exploit rhythmic patterns, since the characteristic rhythm of French is syllable-based. Recent results from studies of speech segmentation in a third language, Japanese, confirm the connection between segmentation and speech rhythm. In Japanese, speech rhythm is based on a subsyllabic unit called the *mora* (which can be a vowel, an onset-vowel sequence, or a syllabic coda). Otake et al. (1993) conducted an experiment in Japanese which was directly analogous to the French experiment by Mehler et al. (1981); they compared detection of CV (e.g. *ta-*) and CVC (e.g. *tan-*) targets in Japanese words beginning with open (*tanishi*) versus closed (*tanshi*) syllables. In both words the first mora is the initial CV sequence *ta*; and detection of CV targets was equally fast in both words (had the Japanese subjects been using a syllabic segmentation procedure, the CV targets should have been harder to detect in closed than in open syllables). CVC targets constitute two morae, and correspond to the first two morae of the words with initial closed syllables; however, they do not correspond properly to a mora-based segmentation of words like *tanishi* (CV-CV-CV). Indeed, the Japanese listeners responded to the CVC targets in words like *tanshi*, but usually failed to respond in words like *tanishi*.

Thus rhythmic segmentation seems to be quite a widespread phenomenon across languages, with the nature of the rhythmic processing being determined by the nature of each language's characteristic rhythmic structure: stress-based, syllabic, or moraic rhythm can all be used in speech segmentation by adult listeners. However, there turn out to be strict limitations on the way any listener can exploit speech rhythm in segmentation; and these limitations may illuminate the questions with which we started this chapter, namely those pertaining to how the prelinguistic infant first solves the segmentation problem.

7. Limits to rhythmic segmentation

First of all, the cross-linguistic differences in speech segmentation are characteristics of the listeners and do not simply follow from the nature of the speech input. English monolinguals do not use syllabic segmentation in performing the target detection task with either English or French (Cutler et al. 1986); neither English nor French listeners use moraic segmentation when performing the same task with Japanese (Otake et al. 1993). In other words, syllabic segmentation seems to be specific to French listeners, moraic segmentation to Japanese listeners. (In fact the French listeners segmented both English and Japanese speech by syllables, just as they segment French!)

Moreover, under appropriate conditions listeners can be seen to abandon the rhythmic segmentation procedures characteristic of their language community. When responding very fast, French listeners can base their responses on subsyllabic units (Dupoux 1993). CVV sequences are apparently less conducive to application of moraic segmentation by Japanese listeners than the (more common) CVCV and CVN sequences (Otake 1992). The failure to find processing disadvantages for English words beginning with weak syllables when the words are carefully read, reported in the second section of this paper, may reflect a similar case: if the input is very clear, stress-based segmentation may not need to be called into play. Thus it is quite clear that none of the rhythmic segmentation procedures constitutes an absolutely *necessary* component of adult listeners' speech processing.

The strongest evidence that this is so comes, however, from studies of bilingual processing. Cutler et al. (1992) tested French-English bilinguals with the techniques which had demonstrated syllabic responding in French listeners (Mehler et al. 1981) and stress-based responding in English listeners (Cutler and Norris 1988). Their subjects were as bilingual as they could find – each had learned both languages from the earliest stages of acquisition, spoke both languages daily, and was accepted as a native speaker by monolinguals in each language.

Yet these bilinguals did not necesarily produce the pattern of results which monolinguals had shown on each previous experiment. Instead, their response patterns could be predicted from a measure of what Cutler et al. called language 'dominance', which amounted in essence to a decision as to which of their two languages the bilinguals would be most sorry to lose. On Mehler et al.'s target detection task with French materials, only those bilinguals who chose French as their 'dominant' language showed a syllabic pattern of responding; the English-dominant bilinguals showed no trace of syllabic effects. On Cutler and Norris' word-spotting task, in contrast, a stress-based response pattern appeared only with those bilinguals who chose English; the responses of the French-dominant bilinguals were unaffected by the rhythmic pattern of the embedding nonsense word. Apparently, these maximally competent bilinguals had available to them in these tasks only one rhythmic segmentation procedure – either that which was characteristic of one of their native languages, or that which was characteristic of the other, but not both.

Of course, it should be remembered that this conclusion is based only on the results of laboratory experiments, and may not reflect the full extent of the resources which bilinguals can apply to the processing of, for example, spontaneous speech; as earlier sections of this chapter described, different speech styles may call differentially upon a listener's processing repertoire. However, the experiments undeniably show that in the laboratory some bilinguals can exploit a given rhythmic segmentation procedure, and do exploit it, while others certainly do not exploit the same procedure, and possibly cannot do so. A claim that, for example, French-dominant French-English bilinguals are capable of stress-based segmentation, but abandon it when processing laboratory speech, ought therefore to be accompanied by an account of why English-dominant bilinguals, and monolingual English speakers, do *not* abandon this procedure in the laboratory. On the basis of the laboratory results alone, it would surely appear that bilinguals simply do not have available to them the segmentation procedure characteristic of their non-dominant language.

This is a remarkable finding in the light of the undoubted competence of these bilingual speakers in *both* their languages. The English-dominant bilinguals spoke and understood French just as well as the French-dominant bilinguals did, and the latter group spoke and understood English just as well as the former. For those bilinguals who used stress-based segmentation with English, the apparent unavailability of syllabic segmentation for use with French seemed to have no adverse effect on their linguistic competence; likewise, for those bilinguals who used syllabic segmentation with French, the unavailability of stress-based segmentation seemed not to reduce in any way their demonstrated competence in English. These results may therefore indicate that the rhythmic segmentation procedures are not a necessary component of a language user's processing mechanism; one can demonstrate native competence without them.

This in turn would imply that rhythmic segmentation procedures are not simply developed in response to experience with the statistical properties of the native language, as the arguments made by, for instance, Cutler and Carter (1987) with respect to stress-based segmentation in English contended. There is no doubt that stress-based segmentation *does* work efficiently with English; but despite having been exposed to English since their earliest years, and despite using English with native competence all their lives, the French-dominant bilinguals nevertheless do not, in the word-spotting experiment, show evidence of segmenting by stress. The question must be posed, therefore, of how the rhythmic segmentation procedures could arise, if it is arguably the case that they may not result automatically from experience with the statistical properties of the native language. A possible answer to this question, proposed by Cutler et al. (1992) and by Cutler and Mehler (1993), is discussed in the next section.

8. Rhythmic segmentation by infants?

Suppose that the rhythmic segmentation procedures used by adult listeners exist not for purposes of adult processing at all, but are simply traces which remain from a period when the segmentation problem dominated the infant's language processing. Perhaps it is precisely the characteristic rhythm of the input language which offers the infant a first toehold into lexical acquisition, by suggesting a possible segmentation of the continuous speech stream into discrete units. In the case of the syllable in French, in fact, just such a model has been put forward by Mehler et al. (1990).

More generally, Cutler and Mehler (1993) have suggested that the infant enters the world already armed with what they have called a 'periodicity bias'. The task of lexical acquisition is primed in that the infant expects that meaning will map to form; the task is made possible by the fact that this expectation is targeted towards a particular kind of form: input which is periodically structured. Speech signals have periodic structure, and for the majority of children speech will be among the most salient forms of input available. In the first few months and even days of life infants prefer to listen to speech rather than to other auditory input (Colombo and Bundy 1981, Glenn et al. 1981).

The contrast between the periodic structure of speech as opposed to random noise is only one level of structure, however; regular periodic structure in speech exists at several levels. At the level of the speech sound, some sounds are relatively 'more periodic' than others – for instance, vowels are relatively steady-state sounds, while consonants are often more transient. As Cutler and Mehler (1993) point out, this could account for the findings that infants acquire language-specific vowel prototypes at about six months of age (Kuhl et al. 1992), well before they acquire the consonantal phonology of their language (Werker and Polka 1993). At a higher level again is the rhythmic structure of language discussed in the preceding sections. It is not unreasonable to propose that at some point during the infant's prelinguistic period this level of rhythmic structure is also attended to, and that the processing that goes on at that point is intimately connected with lexical segmentation – dividing the continuous speech into lexical units. Nor need the rhythmic structure be exclusively expressed in the auditory domain; as Pettito and Marentette (1991) demonstrate, gestural language acquisition by congenitally deaf infants follows a developmental path with noticeable similarities to spoken language acquisition.

Because, as we have described, rhythmic structure differs even across spoken languages, the infant exposed to stress rhythm will focus upon a different regularity than the infant exposed to, say, syllabic or moraic rhythm. As Cutler et al. (1992) argue, this can be conceived of as the infant attending to the smallest level of regularity occurring in the spoken input. What is remarkable about this process is that it seems to happen only once, if the evidence from the bilingual studies is reliable. That is, exposure to two differing rhythmic regularities (syllabic and stress rhythm, for instance) does not result in the ability to use both types of rhythm in speech segmentation; a language user appears to be able to command only one rhythmic segmentation procedure. This type of all-or-none instantiation of a language processing procedure is distinctly reminiscent of the notion of parameter-setting in syntactic processing (e.g. Wexler and Manzini 1987).

9. Prosody and the prelinguistic infant

The notion of rhythmic segmentation by infants as a 'bootstrap' into the beginnings of lexical acquisition has not been directly tested. In this concluding section, however, research on prosodic processing by prelinguistic infants will be reviewed in an attempt to discover whether there is evidence which could support the notions proposed above. First, though, it should be acknowledged that involvement of speech prosody in this level of language acquisition has been proposed by others in several forms. The suggestion by Gleitman and Wanner (1982; see also Gleitman et al. 1987) that words are initially identified as units via their stressed syllables is closely related to the application of the present proposal to languages with stress rhythm, for instance. Likewise, Jusczyk (e.g. 1993) has suggested that prosodic structure is the dimension which infant listeners exploit to accomplish speech segmentation. In a series of experiments Jusczyk and his colleagues have shown that infants are sensitive to prosodic marking of syntactic structure, with sensitivity to clausal prosody emerging by four to five months of age (Hirsch-Pasek et al. 1987), and sensitivity to phrasal prosody by nine months (Jusczyk et al. 1992); Jusczyk interprets this pattern as evidence of infants' exploitation of utterance prosody to structure speech into interpretable units.

Indirect evidence for the present proposal can be found in both perceptual and production evidence from prelinguistic infants. For instance, it has been shown that the characteristic rhythmic pattern of speech is salient to the newborn child. Condon and Sander (1974) found that neonates are able to synchronise their movements with speech structure, whether the speech is spoken directly to the child or played from a tape recorder, and whether it is in the parental language or a foreign language. (Tapping sounds, on the other hand, did not invoke synchrony in the infant's movement.) The ability to discriminate the contrasts involved in rhythmic patterning appears early; thus two-month-olds can discriminate rhythmic groupings of tones (Demany et al. 1977). These early discriminatory abilities also apply to the particular contrasts involved in speech rhythm: very young infants can discriminate stress contrasts (Spring and Dale 1977, Jusczyk and Thompson 1978, Karzon 1985), and neonates can make discriminations based on number of syllables (Bijeljac-Babic et al. 1993). Speech to infants tends to have more regular rhythm than speech to adults, as evidenced in English by more frequent occurrence of stresses (Garnica 1977) and more regular alternation of vocalisation and pause (Stern et al. 1983); however, the relevance of this is unclear given that durational features of infant-directed speech do not appear to be involved in infant preferences for this speech style (Fernald and Kuhl 1987).

More important would seem to be recent evidence of rhythmic patterning in the speech production of prelinguistic infants. Cross-linguistic studies of babbling have pointed to increasing language-specificity in babbling during the second half of the first year of life (e.g. de Boysson-Bardies et al 1984, de Boysson-Bardies and Vihman 1991, Hallé et al. 1991, Blake and de Boysson-Bardies 1992), including language-specificity in prosodic structure (Whalen et al. 1991). Rhythmic structure is one of the language-specific patterns which appear in speech at this age. Levitt and Wang (1991) and Levitt and Utman (1992) found that reduplicative babbling of infants from French-speaking homes showed a gradually increasing regularity of timing of non-final syllables across the first year of life, while the speech of infants of the same age from English-speaking homes showed a gradually increasing variability of syllable structure and timing. This suggests that the characteristic rhythm of speech is incorporated into infants' linguistic competence before they acquire their first words.

It appears that infants also become aware of the characteristic word prosody of their language before acquiring their first words. Jusczyk et al.

(1993) found that nine-month-old infants acquiring English preferred to listen to lists of bisyllabic words with initial stress (*crossing*, *former*, *cable*) than bisyllables with final stress (*across*, *before*, *decay*), although no such preferences appeared with six-month-olds. Even when the lists were low-pass filtered to remove most of the segmental information, nine-month-olds still preferred the initial-stress lists, suggesting that their preferences were based on prosodic structure. Jusczyk et al. argued that during the second half of their first year, infants exercise their ability to segment speech with the result that they acquire knowledge of the typical prosodic structure of words in the input language.

At later ages, language-specific exploitation of rhythmic structure by children is established: children learning English use stress rhythm in segmentation (Gerken 1991, Gerken et al. 1990, Peters 1985); children learning French and other languages with syllable rhythm use syllables (Alegria et al. 1982, Content et al. 1986); children learning Japanese use morae (Mann 1986). The hypothesis proposed here is that language rhythm is also what allows infants to accomplish their very *first* segmentation of speech. An ability to process rhythm is inborn. By using this ability, infants are enabled to overcome the segmentation problem and hence take their first step towards compilation of their very own lexicon.

References

Alegria, J., E. Pignot and J. Morais, 1982. Phonetic analysis of speech and memory codes in beginning readers. Memory and Cognition 10, 451–456.

Bard, E.G. and A. Anderson, 1983. The unintelligibility of speech to children. Journal of Child Language, 10, 265–292.

Bard, E.G. and A. Anderson, 1991. The unintelligibility of speech to children: Effects of referent availability. Proceedings of the Twelfth International Congress of Phonetic Sciences, Aix-en-Provence, Vol. 4, 458–461.

Bard, E.G., R.C. Shillcock and G. Altmann, 1988. The recognition of words after their acoustic offsets in spontaneous speech: Effects of subsequent context. Perception and Psychophysics 44, 395–408.

Barik, H.C., 1977. Cross-linguistic study of temporal characteristics of different types of speech materials. Language and Speech 20, 116–126.

Bernstein-Ratner, N., 1984a. Patterns of vowel modification in mother-child speech. Journal of Child Language 11, 557–578.

Bernstein-Ratner, N., 1984b. Phonological rule usage in mother-child speech. Journal of Phonetics 12, 245–254.

Bernstein-Ratner, N. and C. Pye, 1984. Higher pitch in BT is *not* universal: Acoustic evidence from Quiche Mayan. Journal of Child Language 11, 515–522.

Bijeljac-Babic, R., J. Bertoncini and J. Mehler, 1993. How do 4-day-old infants categorize multisyllabic utterances? Developmental Psychology 29, 711–721.

Blake, J. and B. de Boysson-Bardies, 1992. Patterns in babbling: A cross-linguistic study. Journal of Child Language 19, 51–74.

Blaauw, E., 1991. Phonetic characteristics of spontaneous and read-aloud speech. Proceedings of the ESCA Workshop on Phonetics and Phonology of Speaking Styles, Barcelona, 12.1–12.5.

de Boysson-Bardies, B. and M.M. Vihman, 1991. Adaptation to language: Evidence from babbling and first words in four languages. Language 67, 297–319.

de Boysson-Bardies, B., L. Sagart and C. Durand, 1984. Discernible differences in the babbling of infants according to target language. Journal of Child Language 11, 1–15.

Briscoe, E.J., 1989. Lexical access in connected speech recognition. Proceedings of the 27th Congress, Association for Computational Linguistics, Vancouver, 84–90.

Brown, G., 1977. Listening to spoken English. London: Longman.

Brown, G.D.A., 1984. A frequency count of 190,000 words in the London-Lund corpus of English conversation. Behavior Research Methods, Instrumentation and Computers 16, 502–532.

Brown, R., 1977. Introduction. In: C.E. Snow, C.A. Ferguson (eds.), Talking to children: Language input and acquisition, 1–27. Cambridge: Cambridge University Press.

Cheshire, J., 1982. Variation in an English dialect. Cambridge: Cambridge University Press.

Cole, R.A. and J. Jakimik, 1978. Understanding speech: How words are heard. In: G. Underwood (ed.), Strategies of information processing, 67–116. London: Academic Press.

Colombo, J. and R. Bundy, 1981. A method for the measurement of infant auditory selectivity. Infant Behavior and Development, 4, 219–233.

Coltheart, M., 1981. The MRC psycholinguistic database. Quarterly Journal of Experimental Psychology 33A, 497–505.

Condon, W.S. and L.W. Sander, 1974. Synchrony demonstrated between movement of the neonate and adult speech. Child Development 45, 456–462.

Content, A., R. Kolinsky, J. Morais and P. Bertelson, 1986. Phonetic segmentation in prereaders: Effect of corrective information. Journal of Experimental Child Psychology 42, 49–72.

Crystal, D. and D. Davy, 1969. Investigating English style. London: Longman.

Cutler, A. and S. Butterfield, 1992. Rhythmic cues to speech segmentation: Evidence from juncture misperception. Journal of Memory and Language 31, 218–236.

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

Cutler, A. and J.M. McQueen, in press. The recognition of lexical units in speech. In: B. de Gelder, J. Morais (eds.), From spoken to written language. Cambridge, MA: MIT Press.

Cutler, A. and J. Mehler, 1993. The periodicity bias. Journal of Phonetics 21, 103–108.

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

Cutler, A., J. Mehler, D.G. Norris and J. Segui, 1986. The syllable's differing role in the segmentation of French and English. Journal of Memory and Language 25, 385–400.

Cutler, A., J. Mehler, D. Norris and J. Segui, 1992. The monolingual nature of speech segmentation by bilinguals. Cognitive Psychology 24, 381–410.

Demany, L., B. McKenzie and E. Vurpillot, 1977. Rhythm perception in early infancy. Nature 266, 718–719.

Dupoux, E., 1993. The time course of lexical processing: The syllabic hypothesis revisited. In: G.T.M. Altmann, R.C. Shillcock (eds.), Cognitive models of speech processing: The Sperlonga Meeting II, 81–114. Cambridge MA: MIT Press.

Dupoux, E. and J. Mehler, 1990. Monitoring the lexicon with normal and compressed speech: Frequency effects and the prelexical code. Journal of Memory and Language 29, 316–335.

Ferguson, C., 1977. Baby talk as a simplified register. In: C.E. Snow, C.A. Ferguson (eds.), Talking to children: Language input and acquisition, 209–235. Cambridge: Cambridge University Press.

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

Fernald, A. and P. Kuhl, 1987. Acoustic determinants of infant preference for motherese speech. Infant Behavior and Development 10, 279–293.

Fernald, A. and T. Simon, 1984. Expanded intonation contours in mothers' speech to newborns. Developmental Psychology 20, 104–113.

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

Garnica, O., 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, 63–88. Cambridge: Cambridge University Press.

Gerken, L., 1991. The metrical basis for children's subjectless sentences. Journal of Memory and Language 30, 431–451.

Gerken, L., B. Landau and R.E. Remez, 1990. Function morphemes in young children's speech perception and production. Developmental Psychology 26, 204–216.

Gleitman, L.R. and E. Wanner, 1982. Language acquisition: The state of the state of the art. In: E. Wanner, L.R. Gleitman (eds.), Language acquisition: The state of the art, 3–48. Cambridge: Cambridge University Press.

Gleitman, L.R., H. Gleitman, B. Landau and E. Wanner, 1987. Where learning begins: Initial representations for language learning. In: F. Newmeyer (ed.), The Cambridge linguistic survey, 150–193. Cambridge: Cambridge University Press.

Glenn, S.M., C.C. Cunningham and P.F. Joyce, 1981. A study of auditory preferences in nonhandicapped infants and infants with Down's Syndrome. Child Development 52, 1303–1307.

Grieser, D.L. and P.K. Kuhl, 1988. Maternal speech to infants in a tonal language: Support for universal prosodic features in motherese. Developmental Psychology 24, 14–20.

Grosjean, F., 1985. The recognition of words after their acoustic offset: Evidence and implications. Perception and Psychophysics 38, 299–310.

Hallé, P., B. de Boysson-Bardies and M.M. Vihman, 1991. Beginnings of prosodic organisation: Intonation and duration patterns of disyllables produced by Japanese and French infants. Language and Speech 34, 299–318.

Harrington, J.M., I. Johnson and M. Cooper, 1987. The application of phoneme sequence constraints to word boundary identification in automatic, continuous speech recognition. Proceedings of the First European Conference on Speech Technology, Edinburgh, Vol. 1, 163–167.

Harrington, J.M., G. Watson and M. Cooper, 1989. Word boundary detection in broad class and phoneme strings. Computer Speech and Language 3, 367–382.

Hayes, J.R. and H.H. Clark, 1970. Experiments on the segmentation of an artificial speech analogue. In: J.R. Hayes (Ed.), Cognition and the development of language, 221–234. New York: Wiley.

Heath, S.B., 1983. Ways with words: Language, life and work in communities and classrooms. Cambridge: Cambridge University Press.

Hirsch-Pasek, K., D.G. Kemler Nelson, P.W. Jusczyk, K.W. Cassidy, B. Druss and L. Kennedy, 1987. Clauses are perceptual units for young infants. Cognition 26, 269–286.

Johns-Lewis, C., 1986. Prosodic differentiation of discourse modes. In: C. Johns-Lewis (ed.), Intonation and discourse, 199–219. London: Croom Helm.

Johns-Lewis, C., 1987. The perception of discoruse modes. In: M. Coulthard (ed.), Discussing discourse. University of Birmingham: Discourse Analysis Research Monograph No. 14, 249–271.

Jusczyk, P.W., 1993. How word recognition may evolve from infant speech recognition capacities. In: G.T.M. Altmann, R.C. Shillcock (eds.), Cognitive models of speech processing: The Sperlonga Meeting II, 27–55. Cambridge MA: MIT Press.

Jusczyk, P.W. and E. Thompson, 1978. Perception of a phonetic contrast in multi-syllabic utterances by 2-month-old infants. Perception and Psychophysics 23, 105–109.

Jusczyk, P.W., A. Cutler and N.J. Redanz, 1993. Infants' preference for the predominant stress patterns of English words. Child Development 64, 675–687.

Jusczyk, P.W., D.G. Kemler Nelson, K. Hirsch-Pasek, L. Kennedy, A. Woodward and J. Piwoz, 1992. Perception of acoustic correlates of major phrasal units by young infants. Cognitive Psychology 24, 252–293.

Karzon, R.G., 1985. Discrimination of polysyllabic sequences by one- to four-month-old infants. Journal of Experimental Child Psychology 39, 326–342.

Klatt, D.H., 1989. Review of selected models of speech perception. In: W.D. Marslen-Wilson (ed.), Lexical representation and process, 169–226. Cambridge, MA: MIT Press.

Kowal, S., D. O'Connell, E.A. O'Brien and E.T. Bryant, 1975. Temporal aspects of reading aloud and speaking: Three experiments. American Journal of Psychology 88, 549–569.

Kuhl, P.K., K.A. Williams, F. Lacerda, K.N. Stevens and B. Lindblom, 1992. Linguistic experience alters phonetic perception in infants by six months of age. Science 255, 606–608.

Labov, W., 1972. Sociolinguistic patterns. Philadelphia, PA: University of Pennsylvania Press.

Lamel, L. and V.W. Zue, 1984. Properties of consonant sequences within words and across word boundaries. Proceedings of the 1984 International Conference on Acoustics, Speech and Signal Processing 42.3.1–42.3.4.

Levin, H., Schaffer, C.A. and C. Snow, 1982. The prosodic and paralinguistic features of reading and telling stories. Language and Speech 25, 43–54.

Levitt, A.G. and J.G.A. Utman, 1992. From babbling towards the sound systems of English and French: A longitudinal two-case study. Journal of Child Language 19, 19–49.

Levitt, A.G. and Q. Wang, 1991. Evidence for language-specific rhythmic influences in the reduplicative babbling of French- and English-learning infants. Language and Speech 34, 235–249.

Lieberman, P., 1963. Some effects of semantic and grammatical context on the production and perception of speech. Language and Speech 6, 172–187.

Luce, P.A., 1986. A computational analysis of uniqueness points in auditory word recognition. Perception and Psychophysics 39, 155–158.

Mann, V.A., 1986. Phonological awareness: The role of reading experience. Cognition 24, 65–92.

McAllister, J., 1991. The processing of lexically stressed syllables in read and spontaneous speech. Language and Speech 34, 1–26.

McQueen, J.M. and A. Cutler, 1992. Words within words: Lexical statistics and lexical access. Proceedings of the Second International Conference on Spoken Language Processing, Banff, Canada, Vol. 1, 221–224.

Mehler, J., 1981. The role of syllables in speech processing. Philosophical Transactions of the Royal Society, B295, 333–352.

Mehler, J., E. Dupoux and J. Segui, 1990. Constraining models of lexical access: The onset of word recognition. In: G.T.M. Altmann (ed.), Cognitive models of speech processing: Psycholinguistic and computational perspectives, 236–262. Cambridge, MA: MIT Press.

Mehler, J., J.-Y. Dommergues, U. Frauenfelder and J. Segui, 1981. The syllable's role in speech segmentation. Journal of Verbal Learning and Verbal Behavior 20, 298–305.

Mehta, G. and A. Cutler, 1988. Detection of target phonemes in spontaneous and read speech. Language and Speech 31, 135–156.

Miller, G.A., 1967. The psychology of communication. London: Penguin.

Milroy, L., 1980. Language and social networks. Oxford: Blackwell.

Ohala, J.J., 1983. Cross-language use of pitch: An ethological view. Phonetica 40, 1–18.

Ohala, J.J., 1984. An ethological perspective on common cross-language utilization of F0 of voice. Phonetica 41, 1–16.

Otake, T., 1992. Morae and syllables in the segmentation of Japanese. Paper presented to the XXV International Congress of Psychology, Brussels, July.

Otake, T., G. Hatano, A. Cutler and J. Mehler, 1993. Mora or syllable? Speech segmentation in Japanese. Journal of Memory and Language 32, 358–378.

Peters, A.M., 1985. Language segmentation: Operating principles for the perception and analysis of language. In: D.I Slobin (ed.), The crosslinguistic study of language acquisition, Vol. 2: Theoretical issues, 1029–1067. Hillsdale, NJ: Erlbaum.

Pettito, L.A. and P.F. Marentette, 1991. Babbling in the manual mode: Evidence for the ontogeny of language. Science 251, 1493–1496.

Pilon, R., 1981. Segmentation of speech in a foreign language. Journal of Psycholinguistic Research 10, 113–122.

Remez, R.E., P.E. Rubin and S. Ball, 1985. Sentence intonation in spontaneous utterances and fluently spoken text. Paper presented to the Acoustical Society of America, 109th Meeting, Austin, Texas, April.

Schieffelin, B.B., 1985. The acquisition of Kaluli. In: D.I. Slobin (ed.), The crosslinguistic study of language acquisition, Vol. 1: The data, 525–593. Hillsdale, NJ: Erlbaum.

Schieffelin, B.B. and E. Ochs, 1983. A cultural perspective on the transition from prelinguistic to linguistic communication. In: R.M. Golinkoff (ed.), The transition from prelinguistic to linguistic communication, 115–131. London: Erlbaum.

Segui, J., 1984. The syllable: A basic perceptual unit in speech processing. In: H. Bouma, D.G. Bouwhuis (eds.), Attention and performance, Vol. 10, 165–181. Hillsdale, NJ: Erlbaum.

Segui, J., U. Frauenfelder and J. Mehler, 1981. Phoneme monitoring, syllable monitoring and lexical access. British Journal of Psychology 72, 471–477.

Shillcock, R.C., E.G. Bard and F. Spensley, 1988. Some prosodic effects on human word recognition in continuous speech. Proceedings of SPEECH '88 (Seventh Symposium of the Federation of Acoustic Societies of Europe), Edinburgh, 819–826.

Shipman, D.W. and V.W. Zue, 1982. Properties of large lexicons: Implications for advanced isolated word recognition systems. Proceedings of the 1982 International Conference on Acoustics, Speech and Signal Processing, Paris, 546–549.

Shockey, L. and Z.S. Bond, 1980. Phonological processes in speech addressed to children. Phonetica 37, 267–274.

Shute, H.B., 1987. Vocal pitch in motherese. Educational Psychology 7, 187–205.

Spring, D.R. and P.S. Dale, 1977. Discrimination of linguistic stress in early infancy. Journal of Speech and Hearing Research 20, 224–232.

Stern, D.N. S. Spieker, R.K. Barnett and K. MacKain, 1983. The prosody of maternal speech: Infant age and context-related changes. Journal of Child Language 10, 1–15.

Stoel-Gammon, C., 1984. Phonological variability in mother-child speech. Phonetica 41, 208–214.

Sullivan, J.W. and F.D. Horowitz, 1983. The effects of intonation on infant attention: The role of the rising intonation contour. Journal of Child Language 10, 521–534.

Svartvik, J. and R. Quirk, 1980. A corpus of English conversation. Lund: Gleerup.

Tuaycharoen, P., 1978. The babbling of a Thai baby: Echoes and responses to the sounds made by adults. In: N. Waterson, C. Snow (eds.), The development of communication, 111–125. Chichester: Wiley.

Wakefield, J.A., E.B. Doughtie and B.-H.L. Yom, 1974. The identification of structural components in an unknown language. Journal of Psycholinguistic Research 3, 261–269.

Werker, J.F. and L. Polka, 1993. Developmental changes in speech perception: New challenges and new directions. Journal of Phonetics 21, 83–101.

Wexler, K. and R. Manzini, 1987. Parameters and learnability in binding theory. In: T. Roeper, E. Williams (eds.), Parameter setting, 41–76. Dordrecht: Reidel.

Whalen, D.H., A.G. Levitt and Q. Wang, 1991. Intonational differences between the reduplicative babbling of French- and English-learning infants. Journal of Child Language 18, 501–516.

Wilson, M.D., 1988. MRC psycholinguistic database: Machine-usable dictionary, Version 2.0. Behavior Research Methods, Instrumentation and Computers 20, 6–10.