Assessing the Relevance of Prosodic and Phonotactic Cues on Parsing the Speech Stream by Young Language-learners

Abstract

This is a study about how one-year-old Swedish-learning infants presumably use probabilistic information, such as prosody and phonotactic regularity, in segmentation of speech. The variables studied were the Swedish tonal word accents I & II and the distributional regularities of within-word and between-word consonant clusters in Swedish infant-directed speech. The results – which were not as clear-cut as the results obtained in earlier experiments on English-learning infants – suggest that 12-month old Swedish infants might be sensitive to prosodic cues to word boundaries: in experiment 1, altering the phonotactics of the stimuli reversed the infants’ preference for word accent types. However this was not confirmed in experiment 2, instead there was a general preference for listening at the accent II words. The results also suggest that 12-month old Swedish infants might not use phonotactic cues to word boundaries to the extent as expected: in experiment 1 and 2, altering the word accent types did not reverse the infants’ preference for phonotactics. Instead, both in experiment 1 and 2, there was a general preference for listening at the within-word stimuli. When compared with earlier research these findings indicate that infants, besides being able to integrate multiple statistical cues to word boundaries, might early in life be assisted by pattern-recognition in speech segmentation.

Eeva Koponen
Master term paper in Phonetics
January 2001

Supervisor: Francisco Lacerda
Phonetics, Stockholm University

TABLE OF CONTENTS
1. Background

Unlike printed text, fluent speech does not provide systematical spaces or pauses between words. For adults, who already know the language spoken, there are hardly difficulties in segmentation of fluent speech – but how do infants, who yet not have build up a vocabulary, manage to find discernible word units? Infants eventually must first learn language-specific sublexical/physical cues to word boundaries in their ambient language.
1.1 Sublexical cues to English word boundaries

During the second half of the first year, infants develop sensitivity to how stress, phones, and phonemes typically pattern within words (Mattys, Jusczyk, in press). Some regularities in English sound patterns, as indicated below, are: I) allophonic variations, i.e. the phonetic realization of phonemes as a function of their position in words, II) coarticulation, i.e. the phonetic realization of phonemes as a function of the surrounding segments, III) prosody, i.e. the typical stress pattern of words, and IV) probabilistic phonotactics, i.e. the distributional regularities in occurrence of phonemes or sequences of phonemes. How are allophonic variations, coarticulation, prosody and probabilistic phonotactics related to English word boundaries and how can these cues aid infants to parse the speech stream?

I) Allophonic variations sometimes correlate with word boundaries (Bolinger & Gerstam, 1957; Church, 1987; Lehiste, 1960; Umeda & Coker, 1974): for example, the phoneme /t/ is aspirated when it begins a word, but not when it appears in non-initial positions (Mattys & Jusczyk, in press).

II) The degree of coarticulation between neighboring sounds is in part determined by the presence or absence of a word boundary (Johnson, Jusczyk, in press): for example, due to the rounded vowel, there is a high degree of lip rounding when the /k/ in “coo” is produced. However if the /k/ is separated from the rounded vowel by another phoneme such as the /l/ in “clue”, less lip rounding occurs during the production of /k/l. If a word boundary is added between the /k/ and /l/, as in “suck Lou”, even less lip rounding in /k/ is observed. In short, coarticulation within words may play a role in holding the syllables of a word together (Mattys & Jusczyk, in press).

III) Approximately 90% of English content words begin with a strong syllable (Cutler & Carter, 1987). Each stressed syllable marks the onset of a potential word in an utterance: for example, the English bisyllabic trochaic word “doctor” with the strong-weak stress pattern is likely to be parsed more easily than the iambic word “guitar” with the weak-strong stress pattern. However to parse “taris”; i.e. the strong syllable in “guitar is” and the following unstressed monosyllable “is”, is based on segmentation on trochaic footing (Johnson & Jusczyk, in press; Mattys & Jusczyk, in press).

IV) The frequency with which phonemes occur next to each other in natural speech sequences, range from never to very frequently: for instance, a cluster like /zt/ is never found inside words; it is phonotactically illegal in English. A cluster like /sd/ is, though rare, found within words (e.g. “disdain”) whereas /st/ is very frequent (e.g. “stop”, “last” etc.) (Mattys & Jusczyk, in press).

1.2 The probabilistic nature of word segmentation cues

All of the mentioned segmentation cues: allophonic variations, coarticulation, prosody and phonotactics are probabilistic – in the way that infants need to track distributional frequencies of for example, aspirated phonemes in word onsets, of coarticulated phonemes within words, of one sound following another at word boundaries, and of stressed syllables in word onsets –and can therefore be framed as statistical cues (Johnson & Jusczyk, in press).

What about when a statistical segmentation cue is misleading: when there are no allophonic or coarticulatory cues, or when rare (but phonotactically legal) sequences of phonemes are found within words, or when the words have weak-strong stress pattern? Several researchers (Christiansen, Allen, & Seidenberg, 1998; Mattys, Jusczyk, Luce, & Morgan, 1999; Morgan, 1996; Morgan & Saffran 1995; Saffran, Newport, & Aslin, 1996) agree on that “Speech segmentation is a heuristic process whose chances of success increase when cues are combined” (Mattys & Jusczyk, in press). Accordingly, another way of defining the probabilistic nature of these word segmentation cues is to say that they function in a non-deterministic way, i.e. no single cue alone can accurately segment English words. But since the segmentation algorithm seems to integrate multiple cues, which contribute to perceptual cohesiveness, single explicit boundary cues are not necessary.
1.3 Recent speech segmentation research

1.3.1 Research on conflicting prosodic and phonotactic cues
Mattys et al. (1999) attempted to determine if 9-month-olds are differentially sensitive to sequences of consonants that are typical between words compared to those that are typical within words. Further they were interested in how phonotactic and prosodic cues interact and which of the two sets of cues dominate when placed in conflict. To investigate these questions infants were tested with Headturn Preference procedure.

The stimuli used in their experiments were bisyllabic CVC.CVC non-words (a dot indicates a syllabic boundary) with primary stress on either the first or the second syllable (S/W vs. W/S). Stimuli also differed with respect to the phonotactic nature of their cross-syllabic C.C cluster. Clusters had either high probability of occurring inside of words (“within-word” clusters, e.g. “moftuth”) or high probability of occurring at a word juncture (“between-word” cluster, e.g. “moftuth”). They found that:

a) Among non-words with S/W stress pattern the infants preferred within-word clusters over between-word clusters, i.e. they interpreted the S/W CVC.CVC stimuli as one word-like unit.

b) Among non-words with W/S stress pattern the infants preferred between-word clusters over within-word clusters, i.e. they interpreted the W/S CVC.CVC stimuli as two word-like units, the second stressed syllable functioning as a word onset marker.

c) Among non-words with S/W stress pattern, when a 500msec pause was inserted between the two syllables of each stimulus, the infants preferred between-word clusters over within-word clusters (the opposite to condition a), i.e. the pause functioned as a word boundary marker, just as the W/S stress pattern functioned as a word onset marker in condition b.

d) And finally, the infants preferred non-words with S/W stress pattern made of (conflicting) between-word clusters over non-words with W/S stress pattern made of (conflicting) within-word cluster, i.e. the S/W prosodic cue (indicating one word) dominated the phonotactical cue (indicating two words).

1.3.2 Research on phonotactic cues in on-line word segmentation
According to Mattys et al. (1999) their earlier study (discussed in 1.3.1) on conflicting phonotactic and prosodic cues showed that 9-month-olds’ perceptual preferences are consistent with segmentation based on phonotactic regularities, but did not directly demonstrate that infants actually use this sensitivity to segment words from fluent speech. In a more recent study Mattys & Juszcyk (in press) investigated how 9-month-olds make use of phonotactic cues in on-line word segmentation. The goodness of the phonotactic cues was estimated by the frequency with which the C.C clusters at the onset and offset of a CVC test stimulus (i.e. C.CVC.C) are found between words in child-directed speech, with high between-word probability associated with good cues to word boundaries.

Again using the Headturn Preference procedure, they found that infants listened to CVC stimuli longer when the stimulus during habituation appeared in a sentential context with good phonotactic cues than when it appeared in one without such cues. It is also worth noting that the infants were able to segment words from the passages despite the fact that the target words never occurred in sentence-final positions. In addition their investigation showed that similar segmentation results emerged when good phonotactic cues occurred only at the onset (i.e. C.CVC.C) or only at the offset (i.e. C.CVC.C) of the target words in the utterances.
1.3.3 Research on conflicting prosodic/coarticulatory and phonotactic cues

Johnson & Jusczyk (in press) investigated the issue of multiple cue integration by 8-month-olds. Using the Headturn Preference procedure they carried out three experiments: a) they replicated Saffran et al.’s study (1996), which demonstrated infants’ ability to segment a continuous speech stream based on statistical cues alone, b) they pitted stress against statistics, and c) they pitted coarticulation against statistics. Johnson & Jusczyk defined the phonotactical cues as statistical cues, opposed to the prosodic and coarticulation cues defined as speech cues. The stimuli used in their experiments consisted of a 2-minute speech stream concatenated of four tri-syllabic words (“padiku”, “tibudo”, “golatu”, “daropi”). Each of the 12 CV syllables used to create the four words was recorded individually to avoid coarticulation between syllables. The experiments a, b, and c, of Johnson & Jusczyk, differed with respect to the nature of the segmentation cues in the following way:

a) The only cue to word boundaries was the statistical nature of the stimuli: the transitional probability between syllables within a word was always equal to 1 (no syllable appears in more than one word), and the transitional probability between syllables spanning over word boundaries was always equal to one third.

b) Each time a part-word test item (last syllable of one word plus the first two of another) occurred during the habituation its first syllable was replaced with a stressed version 2, i.e. the stress cue suggested one potential word boundary, while statistical cues suggested another.

c) Each time a part-word test item occurred during the habituation it was replaced by new coarticulated version of the part-word 3, i.e. the potential word boundary cues were once again in conflict.

Their (Johnson & Jusczyk, in press) results of experiments a, b, and c showed that:

a) The replication of Saffran et al.’s study (Johnson & Jusczyk made a natural speech version of Saffran et al.’s artificial language) was successful, indicating evidence that infants employ the same general statistical learning mechanism to parse both speech and non-speech stimuli.

b) The stress cues overwhelmed the phonotactical cues.

c) The coarticulatory cues overwhelmed the phonotactical cues - b) and c) together indicating that infants relied more on the speech cues, i.e. stress and coarticulatory cues to word boundaries than they did on the statistical cue, i.e. transitional probabilities of successive syllables.

1 In the passage containing good phonotactic cues for e.g. the non-word “tove”, the stimulus involved the between-word clusters /vt/ at the onset and /vt/ at the offset, as in “…brave tove trusts…” and in the passage without such cues for e.g. the non-word “love”, the stimulus involved the within-word clusters /lt/ at the onset and /vn/ at the offset, as in “…gruff tove knows…”.

2 An orthographic representation of the stressed speech stream:
“tibudogolaTUDaropitibudodaroP1golatu…” (capitalization indicates stressed syllables)

3 An orthographic representation of the coarticulated speech stream:
“tibudogolatubudodaropitibudodaropigolatu…” (underlined syllables signal coarticulated part-words)

1.3.4 Summary – segmentation cues at different ages

Jusczyk, Friederici, Wessels, and Svenkerud (1993), found that 9-month-old American infants listened longer to non-words with phonemic sequences legal in English, but illegal in Dutch than to non-words with sequences legal in Dutch but illegal in English. Dutch infants showed the opposite pattern of preference. In contrast 6-month-olds listened equally to words with legal or illegal sequences. This age breakdown also held when the question was not about the components’ phonotactic legality, but probability. In line with this finding, Mattys et al. (1999) (discussed in 1.3.1) found that 9-month-olds were sensitive to whether sequences of consonants
were typical between words or typical within words.

Another finding, of Mattys et al. (1999), was that when the phonotactic and prosodic cues were put in conflict, 9 month-old infants relied on prosodic regularity only. The authors speculated that the difference between the two cues arises because prosodic segmentation, which appears to be an early and coarse first-pass strategy, is still dominant. Mattys & Jusczyk (in press), (discussed in 1.3.2), further showed that 9-month-olds make use of good phonotactic cues (regardless of them being at the onset and offset, or only at the onset/offset of the target words) in on-line word segmentation. This confirmed, according to the authors, that at 9 months, on-line word segmentation based on phonotactics, has started to emerge.

The experiments of Johnson & Jusczyk (in press) (discussed in 1.3.3) showed that 8-month-olds segmented a continuous natural speech stream based on statistical cues alone, and that speech cues (i.e. stress and coarticulatory cues) dominated over statistical (i.e. phonotactical) cues, when these were put in conflict. Since it has been earlier suggested that infants are sensitive to the rhythmic properties of language right from birth (De Casper & Spence, 1986; Mehler et al., 1988; Nazzi et al., 1998), it was not surprising, according to Johnson & Jusczyk (in press), that English-learners rely on stress cues when beginning to segment words.

And finally, in line with these experiments Morgan & Saffran (1995) showed that 6-month-olds performed equally well with syllables that did or did not have fixed ordering patterns (according to the click detection paradigm), and that 9-month-olds performed best when presented with sequences of syllables that had both fixed ordering and fixed rhythmic pattern. Morgan & Saffran concluded that, by 6 months infants are sensitive to the rhythmic properties of the input only, by 8 months they are sensitive to the statistical properties of the input (Saffran et al., 1996), and by 9 months infants presumably integrate these cues.

2. Rationale

The cues that are useful in segmenting speech sometimes depend on the nature of the sound organization of the language, i.e. the cues are language specific. There are many examples of language specific allophonic variations and language specific phonotactics, but also variations in coarticulation - e.g. rounding differences being more important in Swedish than in English vowel system (Johnson & Jusczyk, in press), and variations in stress patterns across different languages, e.g. languages without a trochaic rhythmic organization, such as French – might serve as a cue for infants to recognize the phonetic distinctions that are meaningful for their native language. Support for use of language specific segmentation cues comes from studies on adults: the segmentation strategy appropriate for the adults native language has been shown to be imposed also on a foreign language (Cutler, Mehler, Norris, & Segui, 1986) as well as on an artificial language (Vroomen, Tuomainen, & de Gelder, 1998).

Thus, because recent speech segmentation work has almost solely been focused on English there is a considerable risk that the results may reflect language specific aspects of phonotactics and prosody. Therefore, it is of importance to study how Swedish prosodic and phonotactic cues may help Swedish-learning infants in their segmentation task.

Of the Swedish tonal word accents: accent I and accent II, accent II is sometimes called “the accent of compounds” (Swe. sammansättningsaccent). This is probably due to the fact that a majority of Swedish compounds are accent II words, irrespective of the lexemes of the compound being separately pronounced as accent I or accent II words (Bruce, 1998). It is also true that, in some Southern Swedish and Northern Swedish dialects (or due to idiolects or
different speaking styles) compounds sometimes are accent I words. There are even some exceptions to the dominance of accent II compounds in Central Swedish, such as names of the days ("måndag", "tisdag" etc), some proper names ("Bergman", "Lundgren"), and a few lexicalised words ("riksdag", "verkstad", "trädgård", "la(du)gård").

My assumption is that the two-peaked fundamental frequency ($f_0$) correlate of the word accent II might function as a useful cue to word boundaries for Swedish-learning infants: just as coarticulation within words may hold the syllables of a word together, the accent II $f_0$ contour of Swedish compounds might connect the two lexemes of a compound together and thereby indicate that the lexemes are part of one single word.

Yet it appears that the sensitivity to certain segmentation cues emerges early, followed soon thereafter by sensitivity to additional cues to word boundaries – the nature of segmentation strategies at different ages is not fully explored. For example – if segmentation of speech by 6 months is based on prosodic cues only, and by 9 months on prosodic cues integrated with phonotactic cues – which set of cues dominate after another three months, i.e. by the approximate age of first word production? To explore segmentation strategies at the dawn of word production, the subjects in the present study were approximately 12 months old.

3. Aim

The aim of this study is to examine if one-year-old Swedish-learning infants are more sensitive to prosodic or probabilistic phonotactic information to word boundaries. Under the assumption that infants in their segmentation task are guided by sublexical cues to word boundaries, this study will address how the Swedish tonal word accents: accent I and II, and distributional regularities in occurrence of consonant clusters within and between words, may provide the infant with segmentation cues on parsing the speech stream.

The general hypothesis is that infants integrate multiple sources of statistical information to discover word boundaries and that by 12 months they might be more sensitive to phonotactic than to prosodic cues.

4. Method – EXPERIMENT 1

To investigate the role of prosodic and probabilistic phonotactics in cuing word boundaries in Swedish, naturally produced CVC.CVC non-words were created.

The word accent of the non-words was either accent II (CV\_C.CVC), or a concatenation of two accent I CVC monosyllables (CV\_C.CV\_C). The $f_0$ contour of accent II was expected to "hold" the CVC.CVC syllables together and thereby indicate a single word-like unit – as
opposed to the two concatenated accent I CVC monosyllables, which were expected to indicate two separate units.

The cross-syllabic C.C clusters varied in the frequency with which phonemes tend to occur next to each other within words and between words in natural infant-directed speech sequences. The C.C clusters with high probability for occurring within-words (e.g. “moskut” /mɔskUt/, “mobrut” /mu:brUt/) were expected to indicate one word-like unit. The C.C clusters with high probability for occurring between-words (e.g. “mondut” /mɔndUt/, “monfut” /mu:nfUt/) were expected to indicate two word-like units.

The type of word accents and phonotactics of the CVC.CVC non-words, shown in Table 4.1, were combined as follows: the accent II CVC.CVC stimuli were made of within-word phonotactics in agreement (i.e. both the prosody and the phonotactics indicated one word-like unit) or of between-word phonotactics in conflict (i.e. the prosody indicated one word-like unit, and the phonotactics indicated two word-like units), and the accent I + accent I CVC.CVC stimuli were made of within-word phonotactics in conflict (i.e. the prosody indicated two word-like units, and the phonotactics indicated one word-like unit) or of between-word phonotactics in agreement (i.e. both the prosody and the phonotactics indicated two word-like units).

The hypothesis was that there would be a difference in infants’ preference of the set of CVC.CVC stimuli in agreement and the set of CVC.CVC stimuli in conflict. The prosodically and phonotactically well formed CVC.CVC stimuli in agreement were thought of sounding cohesive, as opposed to the non-cohesive CVC.CVC stimuli in conflict.

Table 4.1: The prosody and the phonotactics of the naturally produced CVC.CVC non-words were either in agreement, i.e. both of them indicated one word-like unit or both of them indicated two word-like units, or in conflict, i.e. one of them indicated one word-like unit, and the other two word-like units.

<table>
<thead>
<tr>
<th>within-word phonotactics (one unit)</th>
<th>accent II (one unit)</th>
<th>accent I + accent I (two units)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>agreement</td>
<td>conflict</td>
</tr>
<tr>
<td>between-word phonotactics (two units)</td>
<td>conflict</td>
<td>agreement</td>
</tr>
</tbody>
</table>

4.1 Subjects

19 subjects (11 females, and 8 males), (mean age 372.5 days, age range 356 to 392 days) participated in experiment 1. A total of 24 video recordings of the subjects were made: 5 of them participated twice, and 14 participated once in different test conditions of experiment 1. One additional subject was tested but excluded due to interrupted recording (infant crying). The subjects were all from Swedish-speaking homes - in addition English was spoken by one of the mothers to her infant.

To obtain subjects, a letter with a short description of the experiments was sent to 50 parents – to approximately 12-month-old infants – (in or around the city of Stockholm) randomly drawn from the National Swedish Address Register (SPAR). A total of 25 parents volunteered to participate in the experiments (experiment 1 or 2).

4.2 Stimulus materials

The materials consisted of 9 within-word CVC.CVC stimuli: “moskut”, “mostüt”, “mōntü”, “mobrut”, “mōlkt”, “mōgtü”, “mōngtüt” (mɔŋ’tUt), “mōndtū” (mɔnd’tUt) and “mōngtū” (mɔŋ’tUt), and of 9 between-word CVC.CVC stimuli: “mondū”, “monsū”, “mōntū”, “mōngtū”, “mōngtū”, “mōngtū” (mɔng’tUt), “mōrdū” (mɔrd’tUt) and “mōntrū” (mɔntr’Ut). The length of each vowel was either short or long, giving four different length combinations for each CVC.CVC stimulus: short-short (CVC.CVC), long-short (CV.CVC), short-long (CVC.CV:C), and long-long (CV:C.CV:C). Each audio file contained 24 (6 x 4, or 3 x 2 repetitions x 4) randomly ordered within-word or between-word CVC.CVC stimuli. The inter-stimulus-interval (ISI) was 2 seconds, and the
The frequency and the relative frequency (%) of 15 most frequently occurring within-word clusters in the SWIDS corpus are shown in Table 4.2, to the left. Of these clusters: “sk”, “st”, “nt”, “br”, “lk”, “tr”, “ng” (\(\text{C.C}\)), and “rn” (\(\text{C.C.C}\)) were chosen as within-word C.C clusters for the CVC.CVC stimuli. The geminated consonants “tt”, “ll”, “nn”, “mm”, “dd”, “ck” and “pp” were excluded as stimulus materials, thereby avoiding that the length of geminated consonants might effect the interpretation of one/two units.

The frequency and the relative frequency (%) of 15 most frequently occurring between-word clusters in the SWIDS corpus are shown in Table 4.2, to the right. Of these clusters: “nd”, “ns”, “nh”, “nv”, “nk”, “nf”, “rd” (\(\text{r/l}\)), and “rn” (\(\text{r/l}\)) were chosen as between-word C.C clusters for the CVC.CVC stimuli. To match the number of within-word clusters, the remaining between-word clusters “nm”, “j\(h\)”, “nl”, “t\(d\)”, “r\(h\)”, “r\(v\)”, and “t\(v\)” were excluded as stimulus materials.

The clusters “ng”, “rd” and “rn” were used as stimulus materials in a separate test condition of experiment 1. The pronunciation of these clusters was chosen to be \(\text{C.C.C}\) in the within-word condition, and \(\text{C.C.C}\) in the between-word condition. The clusters “ng”, “rd” and “rn” can, besides within-words, across word-boundaries be pronounced as \(\text{C.C.C}\), \(\text{C.C.C}\), \(\text{C.C.C}\), but since they hardly never within-words are pronounced as \(\text{C.C.C}\), \(\text{C.C.C}\), \(\text{C.C.C}\), the clusters pronounced as \(\text{C.C.C}\), \(\text{C.C.C}\), \(\text{C.C.C}\) were regarded as representative within-word clusters, and the clusters pronounced as \(\text{C.C.C}\), \(\text{C.C.C}\), \(\text{C.C.C}\) were regarded as representative between-word clusters.

All of the chosen within-word clusters are of course legal as between-word clusters, as well as all of the between-word clusters are legal as within-word clusters at the C.C lexeme boundary of compounds (but some of the between-word clusters are illegal as initial or final within-word clusters). In experiment 1, the within-word clusters’ corresponding frequency between-words, and the between-word clusters’ corresponding frequency within-words, were not taken to account. In sum: the C.C clusters of CVC.CVC stimuli were chosen on the basis of their high within-word, respectively high between-word frequency of occurrence only (with the exception of excluding geminated within-word consonant clusters, and special criteria for choosing “ng”, “rn” and “rd”). Due to the lexical nature of infant-directed speech, the within-word and the between-word clusters’ frequencies in the SWIDS corpus are probably not directly comparable to those in Swedish adult-directed speech. For example \(\text{r/l}\) and \(\text{r/l}\) were frequent between-word clusters in the corpus because of numerous occurrences of “har du sett?” (Eng. “Have you seen?”) or “den där” (Eng. “that one”). Also due to the size of the corpus some frequent clusters in Swedish adult-directed speech did not occur in the infant-directed speech corpus.

### 4.2.2 Prosody & Coarticulation

To prepare two different types of prosodic cues to word boundaries accent II stimuli (CV.CVC) and accent I + I (CV.C.CV.CVC) stimuli were created. All the CVC.CVC stimuli were controlled for coarticulation between the syllables – each of the accent I CVC syllables was individually recorded (thereby avoiding coarticulation), and each of the accent II CVC syllables was recorded in controlled coarticulatory contexts. More explicitly, to prepare accent I + I (CV.C.CV.CVC) stimuli, each CVC syllable was recorded (in a carrier phrase) as accent I monosyllable, and then simply concatenated (with no pause in between them) with another accent I CVC monosyllable. To prepare accent II stimuli (CV.C.CV.CVC), shown in Table 4.3, each of the first CVC syllables (CV.C.CV.CVC) was recorded in combinations: CV.C.put, CV.C.tut, and CV.C.kut – and each of the second CVC syllables (CV.C.CV.CVC) was recorded in combinations: mo.t.p.CVC, mo.t.a.CVC, mo.t.k.CVC. Then carefully keeping the \(i_0\) contour of accent II intact, each of the first CVC syllables was concatenated with each of the second CVC syllables in sum: the coarticulatory cues for each C.C boundary could be consistent in place of articulation: CVC.p.p.CVC, CVC.t.t.CVC, CVC.k.k.CVC; or inconsistent in place of articulation: CVC.p.t.CVC, CVC.t.p.CVC, CVC.p.k.CVC, CVC.k.p.CVC, CVC.t.k.CVC, CVC.k.t.CVC.

### 4.2.3 Stress

The role of word accents and stress (Swe. “betoning”), on indicating word boundaries, are hard to separate from each other. The CVC.CVC non-words were stressed as follows: the accent I + I stimuli had primary stress on the
first CVC syllable, and primary stress on the second CVC syllable (\textasciicircum{CVC} \textperiodcentered \textperiodcentered \textunderscore \textunderscore \textunderscore \textunderscore \textunderscore\textasciicircum{CVC}), and the accent II stimuli had primary stress on the first CVC syllable, and secondary stress on the second CVC syllable (\textasciicircum{CVC} \textperiodcentered \textunderscore \textunderscore \textunderscore \textunderscore \textunderscore\textasciicircum{CVC}).

Table 4.2: The frequency and relative frequency (%) of occurrence for within-word clusters in the SWIDS corpus are shown to the left, and the frequency and relative frequency (%) of occurrence for between-word clusters in the SWIDS corpus are shown to the right. The shaded rows show the C.C clusters chosen as stimulus materials for test conditions 1, 2, 3, and 4, and the clusters marked with (*) show the C.C clusters chosen as stimulus materials for test condition 5.

<table>
<thead>
<tr>
<th>within-word cluster</th>
<th>within-word frequency</th>
<th>percentage</th>
<th>between-word cluster</th>
<th>between-word frequency</th>
<th>percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>tt</td>
<td>202</td>
<td>15.26%</td>
<td>rd</td>
<td>113</td>
<td>7.90%</td>
</tr>
<tr>
<td>sk</td>
<td>96</td>
<td>7.25%</td>
<td>nd</td>
<td>62</td>
<td>4.34%</td>
</tr>
<tr>
<td>*ng</td>
<td>91</td>
<td>6.87%</td>
<td>ns</td>
<td>56</td>
<td>3.92%</td>
</tr>
<tr>
<td>ll</td>
<td>79</td>
<td>5.97%</td>
<td>nh</td>
<td>47</td>
<td>3.29%</td>
</tr>
<tr>
<td>nn</td>
<td>63</td>
<td>4.76%</td>
<td>nv</td>
<td>47</td>
<td>3.29%</td>
</tr>
<tr>
<td>st</td>
<td>59</td>
<td>4.46%</td>
<td>nk</td>
<td>43</td>
<td>3.01%</td>
</tr>
<tr>
<td>nt</td>
<td>55</td>
<td>4.15%</td>
<td>nf</td>
<td>39</td>
<td>2.73%</td>
</tr>
<tr>
<td>mm</td>
<td>40</td>
<td>3.02%</td>
<td>nm</td>
<td>35</td>
<td>2.45%</td>
</tr>
<tr>
<td>br</td>
<td>38</td>
<td>2.87%</td>
<td>jh</td>
<td>32</td>
<td>2.24%</td>
</tr>
<tr>
<td>lk</td>
<td>34</td>
<td>2.57%</td>
<td>nl</td>
<td>31</td>
<td>2.17%</td>
</tr>
<tr>
<td>dd</td>
<td>31</td>
<td>2.34%</td>
<td>td</td>
<td>31</td>
<td>2.17%</td>
</tr>
<tr>
<td>ck</td>
<td>27</td>
<td>2.04%</td>
<td>rh</td>
<td>29</td>
<td>2.03%</td>
</tr>
<tr>
<td>*rn</td>
<td>26</td>
<td>1.96%</td>
<td>*rn</td>
<td>29</td>
<td>2.03%</td>
</tr>
<tr>
<td>tr</td>
<td>23</td>
<td>1.74%</td>
<td>tv</td>
<td>29</td>
<td>2.03%</td>
</tr>
<tr>
<td>pp</td>
<td>22</td>
<td>1.66%</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4.3: Concatenation of accent II CV \textperiodcentered \textunderscore \textunderscore \textunderscore \textunderscore \textunderscore\textasciicircum{CVC} non-words: each first CVC syllable was recorded with each of the syllables “put”, “tut”, and “kut”, and each second \textunderscore \textunderscore \textunderscore \textunderscore \textunderscore\textasciicircum{CVC} syllable was recorded with each of the syllables “mop”, “mot”, and “mok”. Then each of the first CVC syllables was concatenated with each of the second \textunderscore \textunderscore \textunderscore \textunderscore \textunderscore\textasciicircum{CVC} syllables. Thereby the coarticulatory cues at the C.C boundaries of the CV \textunderscore \textunderscore \textunderscore \textunderscore \textunderscore\textasciicircum{CVC} stimuli could be consistent in place of articulation (shaded boxes) or inconsistent in place of articulation. Two readings of each CVC syllable were made: one with a short vowel (CVC), and one with a long vowel (CV:C).

<table>
<thead>
<tr>
<th>mop.CVC</th>
<th>mot.CVC</th>
<th>mok.CVC</th>
</tr>
</thead>
<tbody>
<tr>
<td>CVC.put</td>
<td>inconsistent place of articulation</td>
<td>inconsistent place of articulation</td>
</tr>
<tr>
<td>CVC.tut</td>
<td>inconsistent place of articulation</td>
<td>consistent place of articulation</td>
</tr>
<tr>
<td>CVC.kut</td>
<td>inconsistent place of articulation</td>
<td>inconsistent place of articulation</td>
</tr>
</tbody>
</table>

4.3 Procedure

A version of the Preferential Listening Procedure (Fernald, 1985), called Visual Preference Procedure, was used in the present experiment. The parent, with the infant in her/his lap, was seated in a chair approximately one meter away from a TV-monitor. A video camera mounted above the TV-monitor recorded the infant watching the film. To make the parent unable to determine the nature of the stimulus materials, s/he listened to music over tight-fitting PELTOR-HT7A headphones.

To catch the infant’s attention, each film started with a silent still image of a bell for two seconds. Whether or not the infant attended the film, it went on showing the two test objects (either a still image of a blonde girl or a brunette) and played the audios made of CVC.CVC non-words.

The experimenter, located in a booth next door, was able to monitor the section taped. If the infant looked away (from any of the test objects) for consecutive 10 seconds, the experimenter skipped to the next film track showing a new test object and playing its audio. If the infant kept looking at the object, it was shown for the entire duration of the audio (about 80sec). According to the same procedure, both objects were shown twice. Each presentation of the test objects was mediated with the silent still image of the bell for 2 seconds. The experimenter was able to talk to the parent over the headphones (e.g. if the infant got down from the lap of the parent and moved away from camera focus).
The total duration of the films could last from 40 seconds – if the infant looked away for 10 seconds per each test image, to about 5.5 minutes (40 seconds per test image, 2 seconds per the bell image) – if the infant kept looking at all of the images and never looked away for consecutive 10 seconds. The exact looking times (as a measure of attention to the audio) were analyzed frame-by-frame by the experimenter offline. The raw data obtained consisted of stopwatch readings corresponding to every change (front, off) in gaze position of the infant.

4.4 Design

There were 5 test conditions in experiment 1: film1, film2, film3, film4, film5 – which all differed regarding the nature of segmentation cues (prosody indicating either one or two units, and phonotactics indicating either one or two units). An additional version of each of the 5 test conditions was made: film1control, film2control, film3control, film4control, film5control. These “control” films were made of the same objects and the same audios as films 1, 2, 3, 4, and 5, but the presentation order of the objects and the audios was changed. This was done to control sequence effects (fatigue or loss of interest towards the end of films). 24 recordings were made for experiment 1: five recordings were made in each of the test conditions 1 and 2, four recordings were made in each of the test conditions 3 and 4, and finally, six recordings were made in test condition 5.

The objects and audio types of the 5 test conditions in experiment 1 are listed in Table 4.4. As an example, in film1, the first object shown was the blonde. Its audio was made of accent II CVC.CVC stimuli with good cues for within-word phonotactics, i.e. the prosody and the phonotactics were in agreement (both indicated one unit). The second object shown was the brunette. Its audio was made of accent I + I CVC.CVC stimuli with good cues for within-word phonotactics, i.e. the prosody and the phonotactics were in conflict (the prosody indicated one unit, but the phonotactics indicated two units). The third object shown was the blonde again. This time the audio to go with the blonde was the same audio just played with the brunette (the prosody and the phonotactics were in conflict). The last object shown was the brunette again. And finally, the audio to go with the blonde was the same audio played with the blonde, when presented as the first object (the prosody and the phonotactics were in agreement). To summarize the type of test materials and the hypothesis in each test condition in experiment 1:

**Film1 & film1control:**
The phonotactics of CVC.CVC stimuli was kept constant: all CVC.CVC stimuli had a high probability within-word frequency of occurrence. The variable manipulated was prosody: the CVC.CVC stimuli were either cohesive accent II or non-cohesive accent I + I words. The hypothesis was that infants would interpret the CVC.CVC stimulus as one word-like unit.

**Film2 & film2control:**
The phonotactics of CVC.CVC stimuli was kept constant: all CVC.CVC stimuli had a high probability between-word frequency of occurrence. The variable manipulated was prosody: the CVC.CVC stimuli were either cohesive accent I + I or non-cohesive accent II words. The hypothesis was that infants would interpret the CVC.CVC stimulus as two word-like units.

**Film3 & film3control:**
The prosody of CVC.CVC stimuli was kept constant: all CVC.CVC stimuli were accent II words. The variable manipulated was phonotactics: the CVC.CVC stimuli were either made of cohesive within-word phonotactics or of non-cohesive between-word phonotactics. The hypothesis was that infants would interpret the CVC.CVC stimulus as one word-like unit (just like in film1 & film1control).

**Film4 & film4control:**
The prosody of CVC.CVC stimuli was kept constant: all CVC.CVC stimuli were accent I + I words. The variable manipulated was phonotactics: the CVC.CVC stimuli were either made of cohesive between-word phonotactics or of non-cohesive within-word phonotactics. The hypothesis was that infants would interpret the CVC.CVC stimulus as two word-like units (just like in film2 & film2control).

**Film5 & film5control:**
The prosody of CVC.CVC stimuli was kept constant: all CVC.CVC stimuli were accent II words. The variable manipulated was phonotactics: the CVC.CVC stimuli were either made of cohesive within-word C.C boundary clusters /d/, /s/, and /t/, or of non-cohesive between-word C.C boundary clusters /ng/, /rd/, and /rd/. The hypothesis was that infants would interpret the CVC.CVC stimulus as one word-like unit.
Table 4.4: The five different test conditions in experiment 1: film1, film2, film3, film4, and film5 are shown to the left and their “control” films (differing only regarding the presentation order of the objects and the audios): film1control, film2control, film3control, film4control, and film5control are shown to the right. The columns (from left to right) for each film show the order of the objects presented, the type of audio (agreement, or conflict) used, and the phonotactics (within-word or between-word), and prosody (accent II or accent I + I) of the CVC.CVC stimuli. The shaded portions show the variable kept constant throughout each film.

<table>
<thead>
<tr>
<th>film1</th>
<th>film1control</th>
</tr>
</thead>
<tbody>
<tr>
<td>blonde</td>
<td>agreement</td>
</tr>
<tr>
<td>brunette</td>
<td>conflict</td>
</tr>
<tr>
<td>blonde</td>
<td>conflict</td>
</tr>
<tr>
<td>brunette</td>
<td>agreement</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>film2</th>
<th>film2control</th>
</tr>
</thead>
<tbody>
<tr>
<td>blonde</td>
<td>conflict</td>
</tr>
<tr>
<td>brunette</td>
<td>agreement</td>
</tr>
<tr>
<td>blonde</td>
<td>agreement</td>
</tr>
<tr>
<td>brunette</td>
<td>conflict</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>film3</th>
<th>film3control</th>
</tr>
</thead>
<tbody>
<tr>
<td>blonde</td>
<td>agreement</td>
</tr>
<tr>
<td>brunette</td>
<td>conflict</td>
</tr>
<tr>
<td>blonde</td>
<td>conflict</td>
</tr>
<tr>
<td>brunette</td>
<td>agreement</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>film4</th>
<th>film4control</th>
</tr>
</thead>
<tbody>
<tr>
<td>blonde</td>
<td>agreement</td>
</tr>
<tr>
<td>brunette</td>
<td>conflict</td>
</tr>
<tr>
<td>blonde</td>
<td>conflict</td>
</tr>
<tr>
<td>brunette</td>
<td>agreement</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>film5</th>
<th>film5control</th>
</tr>
</thead>
<tbody>
<tr>
<td>blonde</td>
<td>agreement</td>
</tr>
<tr>
<td>brunette</td>
<td>conflict</td>
</tr>
<tr>
<td>blonde</td>
<td>conflict</td>
</tr>
<tr>
<td>brunette</td>
<td>agreement</td>
</tr>
</tbody>
</table>

4.5 Apparatus
The stimuli were recorded by a female native speaker of Swedish, with a BRÜEL & KJÆR 4165 microphone in an anechoic booth. All stimuli were on-line digitized in to a PC at a 41-kHz sampling rate, 16-bit, and then prepared and concatenated, in Cool-Edit-Pro software, into audio lists. Digitized versions of the audio lists, and images for the videos were transferred to Adobe-Premier video-editing software for playback during the experiment. The computer controlled the presentation of the video and audio (except when the experimenter skipped to the next film track if the infant looked away for 10 consecutive seconds). The output was shown via a 24” TV-monitor, and the audio was played through the TV-monitor loudspeaker, at approximately 70dB at the subject’s location. A digital PANASONIC DX110 light-sensitive video camera, mounted just above the TV-monitor, recorded a close-up image of the subject. The test booth 2 x 2.5 meters was dimly lit, and a dark blue curtain suspended around the top of the booth shielded the infant’s view of the rest of the room.

4.6 Evaluation of coarticulatory cues
Since the coarticulatory cues consistent in place of articulation could, in addition to the accent II f0 contour of the CVC.CVC stimuli, indicate one unit (and the coarticulatory cues inconsistent in place of articulation could be in conflict with the accent II f0 contour of the CVC.CVC stimuli), an informal test with one subject (a phonetician) was constructed to decide whether the coarticulatory cues (consistent or inconsistent in place of articulation) were salient.

The subject listened to a total of 144 accent II CV.CVC non-words: 6 within-word, and 6 between-word CVC.CVC non-words, each C.C boundary indicating either consistent place of articulation (CVC.p.p.CVC, CVC.t.t.CVC, CVC.k.k.CVC) or inconsistent place of articulation (CVC.p.t.CVC, CVC.t.p.CVC, CVC.p.k.CVC, CVC.k.p.CVC, CVC.t.k.CVC, CVC.k.t.CVC), plus an additional 36 uncut CVC.CVC “foils” (e.g. “mosput”, “mostut”, “moskut”). To avoid making the test too long, the vowel length of each of the CVC.CVC stimuli was
randomly selected to be short-short (CVC.CVC), long-short (CV:C.CVC), short-long (CVC.CV:C), or long-long (CV:C.CV:C). The subject was asked to respond spontaneously if the CV:C.CVC non-words were/sounded like one word-like unit (CVC.CVC) or like two separate word-like units (CVC and CVC). The test was organized in 12 blocks of 12 stimuli.

The results (Appendix 1) showed that: a) among the concatenated CVC.CVC stimuli consistent in place of articulation (CVC.p-p.CVC, CVC.t-t.CVC, CVC.k-k.CVC) 44% of the stimuli were perceived as “one unit”, and 56% as “two units”, b) among the concatenated stimuli inconsistent in place of articulation (CVC.p-t.CVC, CVC.t-p.CVC, CVC.p-k.CVC, CVC.k-p.CVC, CVC.t-k.CVC, CVC.k-t.CVC) 53% were perceived as “one unit”, and 47% as “two units”, and c) among the uncut “foils” (which naturally all were coarticulated) 53% were perceived as “one unit”, and 47% as “two units”. In sum: the results showed that consistency/inconsistency in place of articulation did not function as a coarticulatory cue to one unit/two units. Due to these results, half of the accent II CVC.CVC stimuli chosen as stimulus materials for experiment 1 were made of C.C boundaries consistent in place of articulation, and half of C.C boundaries inconsistent in place of articulation.

4.7 Adult validation of the stimulus materials

22 students in phonetics at the Department of linguistics, Stockholm university, listened to the audio files of the films to be shown to the infants: 6 subjects listened to the audios of test condition 1, 4 listened to the audios of test condition 2, 4 listened to the audios of test condition 3, 4 listened to the audios of test condition 4, and 4 listened to the audios of test condition 5. The subjects were asked to respond spontaneously if the CVC.CVC non-words were/sounded like one word-like unit (CVC.CVC) or like two separate words-like units (CVC and CVC). Each test was organized in 8 blocks of 6 stimuli.

The results (Appendix 2) showed that: a) among the audios in agreement, when both the phonotactics and the prosody indicated one word, 69% of the stimuli were perceived as “one word”, and 31% as “two words”, b) among the audios in agreement, when both the phonotactics and the prosody indicated two words, 59% of the stimuli were perceived as “one word”, and 41% as “two words”, and c) among the audios in conflict (i.e. when the prosody and the phonotactics of the audio indicated different word boundaries) 52% of the stimuli were perceived in accordance with the one word/two words cues indicated by the prosody of the stimuli, and 48% of the stimuli were perceived in accordance with the one word/two words cues indicated by the phonotactics of the stimuli. In the test, as whole, there was a bias for “one word” responses (59,37% “one word” responses – 40,63% “two words” responses).

One of the subjects’ comments on the stimuli was, that they could tell whether the stimuli were one or two units, but they did not know what exactly affected their interpretation, i.e. the subjects were aware of the stimuli sounding more or less cohesive, but not aware of whether it was due to the prosody or phonotactics or other aspects of the stimuli.

5. Results – EXPERIMENT 1

Mean percentage looking times and their 95% confidence intervals (also shown in Figure 5.1), median percentage looking times, and standard errors of the means in experiment 1 were as follows:

**Film1 & film1control:**
Mean/median 45,64/39,89 (ST.ERR.=7,41) for accent II words, and mean/median 59,54/58,11 (ST.ERR.=3,64) for accent I + I words – i.e. among within-word stimuli (indicating one word) the infants attended longer to accent I + I words in conflict (indicating two words) than to accent II words in agreement (indicating one word).

**Film2 & film2control:**
Mean/median 53,16/55,53 (ST.ERR.=5,36) for accent II words, and mean/median 48,91/47,59 (ST.ERR.=7,16) for accent I + I words – i.e. among between-word stimuli (indicating two words) the infants attended longer to
accent II words in conflict (indicating one word) than to accent I + I words in agreement (indicating two words).

**Film3 & film3control and film5 & film5 control**:  
Mean/median 49.5/49.61 (ST.ERR.=3.29) for between-word stimuli, and mean/median 54.17/54.37 (ST.ERR.=3.37) for within-word stimuli – i.e. among accent II words (indicating one word) the infants attended longer to within-word phonotactics in agreement (indicating one word) than to between-word phonotactics in conflict (indicating two words).

**Film4 & film4control**:  
Mean/median 44.57/45.6 (ST.ERR.=3.6) for between-word stimuli, and mean/median 47.99/52.39 (ST.ERR.=8.25) for within-word stimuli – i.e. among accent I + I words (indicating two words) the infants attended longer to within-word phonotactics in conflict (indicating one word) than to between-word phonotactics in agreement (indicating two words).

*The results of film3 & film3control and film5 & film5control are here collapsed since these conditions did not differ regarding the nature of segmentation cues. Separate figures for these conditions are shown in Appendix 3.*

The within-word condition and the between-word condition (on the top row of **Figure 5.1**) show a slight interaction between the altered phonotactics and preference for word accents: the infants’ listening preference of word accent I + I in the within-word condition was turned around to preference of accent II in the between-word condition.

The accent II condition and the accent I + I condition (on the bottom row of **Figure 5.1**) show no interaction between the altered word accents and preference for phonotactics: the infants’ listening preference of within-word phonotactics was not turned around, but remained the same in the accent I + I condition.

An analysis of variance (ANOVA) performed on these data did not reveal any significant differences or interactions.
6. Discussion – EXPERIMENT 1

When interpreting results from the current experiment a question arises: If the infants perceive, for example the within-word II CVC.CVC stimuli as single word-like units, should they listen longer at the stimuli made of phonotactically better formed within-word clusters (cues in agreement) or phonotactically less well formed between-word clusters (cues in conflict)? The question being difficult to answer, the assumption underlying this discussion is rather that the listening preference of the infants’ – irrespective of whether the cues in agreement or the cues in conflict should lead to longer listening – should be reversed with the opposite phonotactic or prosodic cues to word boundaries.

The results from test conditions 1 and 2 reveal that the infants’ listening preference for word accents was reversed by altering the phonotactic cue: in test condition 1 (within-word condition) the infants preferred the accent I + I stimuli over the accent II stimuli, and in test condition 2 (between-word condition) they preferred the accent II stimuli over the accent I + I stimuli. Thereby one can speculate that the CVC.CVC stimuli were interpreted (in line with the hypothesis) as one word-like units by the infants in test condition 1, and as two word-like units in test condition 2.

The results from test conditions 3&5 and 4 reveal that the infants’ listening preference for phonotactics was not reversed by altering the word accents: in test condition 3&5 (accent II condition) the infants preferred the within-word phonotactics over the between-word phonotactics, and in test condition 4 (accent I + I condition) they again preferred the within-word phonotactics over the between-word phonotactics. Due to preference for within-word phonotactics in both of these conditions, one can speculate that the CVC.CVC stimuli were not interpreted (as expected according to the hypothesis) as one word-like unit by the infants in test condition 3&5, and as two word-like units in test condition 4.

In line with these results, the adult validation of the stimulus materials showed that there was only slightly stronger influence of the prosody (about 52%) than of the phonotactics (about 48%) on signaling word boundaries. Also, neither informal listening nor acoustic analysis of the stimuli confirmed the salience of the word accent contrast – in fact in many cases the f_0 contours of accent II and accent I + I stimuli were almost identical. A reason for the non-prominent within-word/between-word phonotactic contrast (both in the adult validation and in experiment 1 on infants) was probably that the within-word and between-word clusters chosen from the SWIDS corpus shared phonotactical characteristics. Another weakness of the stimuli in experiment 1 was that concatenation of the stimuli sometimes resulted in unnatural C.C boundaries (e.g. slightly shorter or longer pause in between the syllables), which might have influenced the subjects’ interpretation of one unit/two units.

7. Method – EXPERIMENT 2

The purpose of experiment 2 was (just like in experiment 1) to investigate the role of prosodic and probabilistic phonotactics in cuing word boundaries in Swedish. The test materials in experiment 2 were naturally produced CVC.CVCCVC non-words.

The word accent of the non-words was either accent II CVC.CVCCVC (like for the Swedish compound “boʊkkhyllan“, Eng. “bookshelves”), or accent I + II CVC.CVCCVC (like for the Swedish words “deʊn hyˈllan“, Eng. “the shelves”). The concatenated accent I CVC + accent II CV.CCV.CCV syllables were expected, better than the concatenated accent I CV.CCV.CCV syllables in experiment 1, to indicate two separate units. Even this time the f_0 contour of accent II was expected to indicate a single word-like unit.
The cross-syllabic C.C clusters of the CVC.CVCCVC stimuli in experiment 2 were chosen according to the criteria that the frequency of occurrence within-word/between-words in the SWIDS corpus had to be as high as possible, and the corresponding frequency of occurrence between-words/within-words in the SWIDS corpus had to be as low as possible. The C.C clusters with high probability for occurring within-words (e.g. “bettattan” /betːːtːan/, “beskattan” /beskːatːan/), but low probability for occurring between-words, were expected to indicate one word-like unit, and the between-word C.C clusters with high probability for occurring between-words (e.g. “berdattan” /berdːatːan/, “bendattan” /bendːatːan/), but low probability for occurring within-words, were expected to indicate two separate units.

7.1 Subjects
16 subjects (11 females, and 5 males), (mean age 379.5 days, age range 356 to 399 days) participated in experiment 2. A total of 16 video recordings of the subjects were made: 10 of the subjects had participated in experiment 1, and 6 participated for the first time in experiment 2. One additional subject was tested but excluded due to interrupted recording (infant crying). The subjects were all from Swedish-speaking homes – in addition German, respectively Spanish was spoken by one of the mothers to her infant.

7.2 Stimulus materials
The materials consisted of 6 within-word CVC.CVCCVC stimuli: “bettattan”, “beskattan”, “bellattan”, “bengattan”, “begattan”, “bennattan”, and 6 between-word CVC.CVCCVC stimuli: “berdattan”, “bendattan”, “bendattan”, “benhattan”, “benvattan”, “benfattan”. The length of each vowel was always short. Each audio file contained 24 (6 x 4 repetitions) randomly ordered within-word or between-word CVC.CVCCVC stimuli. The ISI was 1 second, and the total duration of each audio was approximately 50 seconds. The reason for the stimuli to appear in a more rapid manner (ISI in experiment 1 was 2 seconds) was to make it easier for the infants to extract the prosodic pattern of the stimuli. The shorter total duration of each audio (total duration of each audio in experiment 1 was approximately 80 seconds) was hoped to lead to skipping to next image only due to the stimuli being perceived as cohesive/non-cohesive (instead of skipping to the next object due to fatigue of looking at an object for the entire duration of the audio).

7.2.1 Phonotactics
The 15 most frequently occurring within-word clusters in the SWIDS corpus, organized according to the criteria that the frequency of within-word occurrence had to be as high as possible, and the corresponding frequency of occurrence between-words had to be as low as possible, are shown in Table 7.1 to the left. Of these clusters: “tt”, “sk”, “lt”, “ng”, “sf”, and “nn” were chosen as within-word C.C clusters for the CVC.CVCCVC stimuli. The remaining clusters were excluded as stimulus materials.

The 15 most frequently occurring between-word clusters in SWIDS, organized according to the criteria that the frequency of between-words occurrence had to be as high as possible, and the corresponding frequency of occurrence within-word had to be as low as possible, are shown in Table 7.1 to the right. Of these clusters: “rd”, “nd”, “ns”, “nh”, “nv”, and “nf” were chosen as between-word C.C clusters for the CVC.CVCCVC stimuli. The remaining clusters were excluded as stimulus materials.

7.2.2 Prosody & Coarticulation
Since the results of the evaluation of coarticulatory cues had shown that consistent place of articulation did not function as a cue to one unit, the CVC syllables of the CVC.CVCCVC stimuli in experiment 2 were not individually recorded or recorded in various consistent/inconsistent place of articulation combinations (as opposed to experiment 1). Instead, to prepare two different types of prosodic cues to word boundaries the accent II CVC.CVCCVC stimuli and the accent I + II CVC.CVCCVC stimuli were simply recorded (in a carrier phrase) as whole sequences.

Table 7.1: The frequency and relative frequency (%) of occurrence for within-word clusters in the SWIDS corpus are shown to the left – the third column shows the within-words’ corresponding between-word frequency. The frequency and relative frequency (%) of occurrence for between-word clusters in the SWIDS corpus are shown to the right – the third column shows the between-words’ corresponding within-word frequency. The C.C clusters chosen as test material for experiment 2 (shaded rows) were the ones with highest frequency of occurrence.
within-word/between-word and the lowest corresponding frequency of occurrence between-word/within-word.

### 7.2.3 Stress
The CVC.CVCCVC stimuli used as stimulus materials were stressed as follows: the accent I + II stimuli had primary stress on the first CVC syllable, primary stress on the second CVC syllable, and secondary stress on the last CVC syllable (CVC:CVCCVC:CVCCVC), and the accent II stimuli had primary stress on the first CVC syllable, unstressed second CVC syllable, and secondary stress on the last CVC syllable (CVC:CVCCVC).  

### 7.3 Procedure
This was the same as in the previous experiment, except that the total duration of the films could last from 40 seconds – if the infant looked away for 10 seconds per each test image, to about 3.5 minutes (50 seconds per test image, 2 seconds per the bell image) – if the infant kept looking at all of the images and never looked away for consecutive 10 seconds.

### 7.4 Design
This was the same as in the previous experiment, except that there was no test condition 5. 16 recordings were made for experiment 1: four recordings were made in test condition 1, five recordings were made in test condition 2, four recordings were made in test condition 3, and three recordings were made in test condition 5.

### 7.5 Apparatus
This was the same as in the previous experiment.

### 8 Results – EXPERIMENT 2
Mean percentage looking times and their 95% confidence intervals (also shown in Figure 8.1), median percentage looking times, and standard errors of the means in experiment 2 were as follows:

#### Film1 & film1control:
Mean/median 73.6/78.56 (ST.ERR.=5.04) for accent II words, and mean/median 66.53/70.39 (ST.ERR.=6.33) for accent I + II words – i.e. among within-word stimuli (indicating one word) the infants attended longer to accent II words in agreement (indicating one word) than to accent I + II words in conflict (indicating two words).

#### Film2 & film2control:
Mean/median 58.97/62.67 (ST.ERR.=7.46) for accent II words, and mean/median 55.15/52.51 (ST.ERR.=4.65) for accent I + II words – i.e. among between-word stimuli (indicating two words) the infants attended longer to accent II words in conflict (indicating one word) than to accent I + II words in agreement (indicating two words).
Film3 & film3control:
Mean/median 63.55/62.61 (ST.ERR.=6.18) for between-word stimuli, and mean/median 64.58/66.6 (ST.ERR.=5.27) for within-word stimuli – i.e. among accent II words (indicating one word) the infants attended longer to within-word phonotactics in agreement (indicating one word) than to between-word phonotactics in conflict (indicating two words).

Film4 & film4control:
Mean/median 57.93/60.36 (ST.ERR.=11.24) for between-word stimuli, and mean/median 58.95/59.13 (ST.ERR.=10.68) for within-word stimuli – i.e. among accent I + II words (indicating two words) the infants attended longer to within-word phonotactics in conflict (indicating one word) than to between-word phonotactics in agreement (indicating two words).

**Figure 8.1: Mean percentage looking time and 95% CI in the Within-word (upper, left), Between-word (upper, right), Accent II (down, left), and concatenated Accent I + II (down, right) conditions, in experiment 2.**

The within-word condition and the between-word condition (on the top row of Figure 8.1) show no interaction between the altered phonotactics and preference for word accents: the infants’ listening preference of word accent II in the within-word condition was not turned around, but remained the same in the between-word condition.

The accent II condition and the accent I + II condition (on the bottom row of Figure 8.1) show no interaction between the altered word accents and preference for phonotactics: the infants’ listening preference for within-word phonotactics was not turned over, but remained the same in the accent I + II condition.

An analysis of variance (ANOVA) performed on these data did not reveal any significant differences or interactions.
9. Discussion – EXPERIMENT 2

The results from test conditions 1 and 2 reveal that the infants’ listening preference for word accents was not reversed by altering the phonotactic cue: in test condition 1 (within-word condition) the infants preferred the accent II stimuli over the accent I + II stimuli, and in test condition 2 (between-word condition) they again preferred the accent II stimuli over the accent I + II stimuli. Due to preference for accent II stimuli in both of these conditions, one can speculate that the CVC.CVCCVC stimuli were not interpreted (as expected according to the hypothesis) as one-word like unit by the infants in test condition 1, and as two word-like units in test condition 2. The preference for accent II words in both conditions rather indicates a biased listening at one word-like units.

The results from test conditions 3 and 4 reveal that the infants’ listening preference for phonotactics was not reversed by altering the word accent: in test condition 3 (accent II condition) the infants preferred the within-word phonotactics over the between-word phonotactics, and in test condition 4 (accent I + II condition) they again preferred the within-word phonotactics over the between-word phonotactics. Just like in experiment 1, one can speculate that the CVC.CVCCVC stimuli were not interpreted (as expected according to the hypothesis) as one word-like unit by the infants in test condition 3, and as two word-like units in test condition 4. The preference for within-word phonotactics in both conditions rather indicates a biased listening at one word-like units.

In experiment 2, the phonotactic contrast was expected to be salient: the high probability within-word C.C clusters had a low probability for occurring between-words (yet the within-word C.C clusters’ corresponding between-word frequency was zero only in two cases out of six) and the high probability between-word C.C clusters had a low probability for occurring within-words (yet the between-word C.C clusters’ corresponding within-word frequency of occurrence was zero only in three cases out of six). Also the word accent contrast was expected to salient and more exaggerated (accent II CVC.CVCCVC stimuli contrasted with accent I + II stimuli) in experiment 2 compared with experiment 1.

Another difference of the stimulus materials in experiment 2 was that the CVC.CVCCVC stimuli were recorded as whole sequences, following that the C.C boundaries were coarticulated and there was no pause in between the syllables. Thus the coarticulatory cues might have functioned as a cue to one word-like unit and biased the subjects’ interpretation of one unit/two units. On the other hand, among the concatenated stimuli in experiment 1 – which rather can be thought of indicating separate units – there was as well biased listening preference for the within-word stimuli indicating one word-like unit.

10. General discussion

Earlier experiments with English-learning infants have shown evidence that infants employ statistical learning mechanisms to parse both speech and non-speech stimuli. By the age of 6 months, segmentation of speech has shown to be based on prosodic cues only, and by 9 months of age on prosodic cues integrated with phonotactic cues. The aim of this study was to examine if prosodic or phonotactic cues dominate the speech segmentation of one-year-old Swedish-learning infants. The general hypothesis was that infants integrate multiple sources of statistical information to discover word boundaries and that by 12 months they might be more sensitive to phonotactical than to prosodic cues. The complexity of the results obtained did not allow straightforward analysis of them. In the within-word condition, along with the between-word condition (in experiment 1) it was shown that prosody could slightly modulate the effect of phonotactic word boundary cues. Thereby 12 month-olds can be argued to be able
to integrate both rhythmic and distributional information to group the input into processing chunks. The fact that the altered phonotactic cues to word boundaries reversed the infants’ looking preference for word accents, while the altered prosodic cues to word boundaries did not reverse the infants’ looking preference for phonotactics, indicates that by 12 months Swedish infants are more sensitive to prosodic cues.

Regarding the stimulus materials used, the results indicate that the difference between the two types of word accents – accent II and accent I + I – was weakly salient in experiment 1, and that the more exaggerated word accent contrast – accent II and accent I + II – led to biased listening preference for the accent II words in experiment 2. Concerning phonotactics, the results indicate that the within-word/between-word contrast (or the components phonotactic probability as opposed to legality) was probably not salient enough neither in experiment 1 or 2 and led to biased listening preference for the within-words in both experiments. The general preference for accent II words made of within-word phonotactics can potentially arise from infants storing information about frequently occurring prosodic and phonotactic patterns of words. In Swedish the speech input to infants is likely to include many accent II words, such as their own names, e.g. “Ulla”, “Pelle”, or nicknames, e.g. (“Jan”) “Janne”, (“Charlott”) “Lotta”, (“Nic”) “Nisse”, as well as child-directed versions of adult words, e.g. (“huund”) “voovve”, (“björn”) “naalle”, (“kafttt”) “kisse”, (“fågel”) “pippo”. Thus the source of infants’ knowledge of prosody and phonotactics may be based on pattern-recognition, deriving from melody and sound patterns of words that they are beginning to store.

Saffran et al. (1996) showed that 8 month-old infants could parse a synthetic speech stream, made of non-syllables with help of probabilistic phonotactic cues to word boundaries only. The artificial stimuli in their experiment were less complex and less speech-like than the naturally produced; prosodically, coarticulatorily and phonotactically highly variable stimuli in the current experiment. In addition, in the type of experiments as Saffran et al., the stimulus materials are typically presented under hermetic experimental conditions allowing optimal opportunity for the infants to extract the neighbouring sounds’ transitional probabilities. In real life transitional probabilities might not be as explicit an in experimental situations. From the ontogenetic perspective the infant can be seen as storing “purposeless” information, auditive or other, and as the memory recourses of the infant fill up, information competes for space and leads to restructuring of the stored information (Lacerda & Lindblom, 1997). In conclusion, the pattern-recognition probably plays role for 12 month-old Swedish-learning infants since they have not received sufficient amount of exposure to their ambient language. Apparently, the pressure imposed by limited available recourses on non-specific saving has not yet become big enough for the phonological structure to emerge.

References
2 Bruce, G. (1998): Allmän och svensk prosodi, Praktisk Lingvistik 16, Institutionen för lingvistik, Lunds Universitet, ISSN 0348-484X.
7 Cutler, A., Mehler, J., Norris, D. & Segui, J. (1986): The syllable’s differing role in the

Appendix 1

The results of the evaluation of coarticulatory cues showed that the C.C boundaries of CVC.CVC stimuli consistent in place of articulation (CVCp-pCVC, CV.Ct-CVC, CV.Ck-kCVC) were perceived as one word-like unit 44.44% of time, and as two word-like units 55.56% of time, and that the C.C boundaries of CVC.CVC stimuli inconsistent in place of articulation (CVCp-CVC, CV.Ct-pCVC, CV.Ck-pCVC, CV.Ck-CVC, CV.Ck-CVC, CV.Ck-CVC) were perceived as one word-like unit 52.78% of time, and as two word-like units 47.22% of time. Among the uncut stimuli (“foils”, which naturally were coarticulated) 52.78% were perceived as one word-like unit, and 47.22% were perceived as two word-like units.
Appendix 2

The results of the adult validation of the stimulus materials showed that: a) among the audios in agreement, when both the phonotactics and the prosody indicated one word, there were 69.34% “one word” responses, and 30.65% “two words” responses, b) among the audios in agreement, when both the phonotactics and the prosody indicated two words, there were 59.37% “one word” responses, and 40.62% “two words” responses, and finally c) among the audios in conflict (i.e. when the prosody and the phonotactics of the audio indicated different word boundaries) 51.89% of the responses (51.89% of the responses) were in accordance with the one word/two-word cues indicated by the prosody of the stimuli (shaded boxes), and 48.11% of the responses were in accordance with the one word/two word cues indicated by the phonotactics of the stimuli (white boxes).

### Appendix 3

Mean percentage looking time in test condition 3: film3 & film3contro (to the left) and test condition 5: film5 & film5contro (to the right) of experiment 1, presented as separate data. The table below show the mean, median, range, and standard error of the mean for each of the conditions.
Within-between comparisons for Accent II:

<table>
<thead>
<tr>
<th></th>
<th>film3 &amp; film3control (exp1)</th>
<th>film5 &amp; film5control (exp1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>88</td>
<td>100</td>
</tr>
<tr>
<td>95% CI</td>
<td>95% CI</td>
<td>95% CI</td>
</tr>
<tr>
<td>percentage</td>
<td>60</td>
<td>50</td>
</tr>
<tr>
<td>looking time</td>
<td>70</td>
<td>80</td>
</tr>
<tr>
<td></td>
<td>90</td>
<td>100</td>
</tr>
</tbody>
</table>

Mean, Median, Range, & ST. ERR.:

<table>
<thead>
<tr>
<th></th>
<th>within</th>
<th>between</th>
<th>within</th>
<th>between</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean</td>
<td>60.37</td>
<td>53.03</td>
<td>50.04</td>
<td>47.14</td>
</tr>
<tr>
<td>median</td>
<td>61.11</td>
<td>50.13</td>
<td>52.03</td>
<td>49.53</td>
</tr>
<tr>
<td>range</td>
<td>46.04</td>
<td>53.93</td>
<td>42.12</td>
<td>46.12</td>
</tr>
<tr>
<td>ST. ERR.</td>
<td>5.15</td>
<td>6.16</td>
<td>4.19</td>
<td>3.72</td>
</tr>
</tbody>
</table>