

Asymmetric processing of lexical tonal contrast in Swedish

VERENA FELDER, ELISABET JÖNSSON-STEINER, AND CARSTEN EULITZ
University of Konstanz, Konstanz, Germany

AND

ADITI LAHIRI
University of Oxford, Oxford, England

Languages such as Swedish use suprasegmental information such as tone, over and above segments, to mark lexical contrast. Theories differ with respect to the abstractness and specification of tone in the mental lexicon. In a forced choice task, we tested Swedish listeners' responses to words with segmentally identical first syllables differing in tonal contours (characterized as Accents 1 and 2). We assumed Accent 1 to be lexically specified for a subset of words and hypothesized that this specification would speed up word accent identification. As was predicted, listeners were fastest in choosing the tonally correct word when the accent was lexically specified. We conclude that the processing of surface tonal contours is governed by their underlying lexical structure with tonal specification.

Languages always distinguish words by different segments. Some also use suprasegmental information, such as stress or tone, to mark contrast. The segmental-lexical level of speech input has been looked into extensively in the psycholinguistic literature, and many authors have raised the issue of nonisomorphism between the speech signal and its lexical representation—that is, the idea that the lexical phonological representation does not necessarily correspond one to one to the speech signal but can be more abstract and less detailed (see Gaskell & Marslen-Wilson, 2001; Gow, 2002; Lahiri & Marslen-Wilson, 1991; Lahiri & Reetz, 2002; Wheeldon & Waksler, 2004). The literature on suprasegmental units is by far sparser and has usually been concerned with either the role of stress cues in word recognition (see Cooper, Cutler, & Wales, 2002; Cutler & van Donselaar, 2001; van Donselaar, Koster, & Cutler, 2005) or a comparison of the relative importance of segmental versus suprasegmental cues (see Connell, 2000; Cutler & Chen, 1997; Cutler & Otake, 1999; Schirmer, Tang, Penney, Gunter, & Chen, 2005; Soto-Faraco, Sebastián-Gallés, & Cutler, 2001; Ye & Connine, 1999). Other researchers have further been concerned with the localization of tone-processing areas in the brain (Gandour, 2006). However, in none of these studies have hypotheses on the mental representation of suprasegmental information been experimentally examined.

In this article, we claim that suprasegmental phonological information is stored in a comparable nonisomorphic way as segmental information—an assumption routinely

made on theoretical grounds (Hyman, 2000). We base our investigation on tonal lexical contrasts. Although tone is used in half of the world's languages to mark lexical contrasts, psycholinguistic studies typically deal with Asian tonal languages (see Gandour, 2006; Schirmer et al., 2005; Sekiguchi & Nakajima, 1999; Yip, 2002). In Germanic languages such as English, German, and Dutch, tonal lexical contrasts are very unusual. Still, Scandinavian languages, particularly Swedish and Norwegian, do differentiate words only on the basis of their tonal properties: Swedish *buren*₁ “the cage,” *buren*₂ “carry—PERFECT PARTICIPLE” (Elert, 1972; Haugen, 1967; Kristoffersen, 2000). In contrast to African and Asian languages, in Swedish and Norwegian, only a few words form minimal pairs based on tone; Elert (1972, 1981) has provided a list of 350 minimal pairs for Swedish. However, minimal pairs are often not numerous for some segmental contrasts either. For instance, [ð] and [θ] are contrastive in English, but there are no minimal pairs. In Scandinavian languages, for the most part, the segmental level is sufficient to determine the correct word, and tonal information is additional information. There are even Swedish and Norwegian dialects that do not make use of tonal information. Nevertheless, standard Swedish is clearly considered to be a tonal language in which tone plays an important role in the morphology–phonology interaction. Since tone is part of the linguistic system, we expect tone to play a decisive role in speech perception. Furthermore, our analyses of morpho-phonological alternations allow clear-cut hypotheses on the representation of these tonal contrasts in the mental lexicon.

V. Felder, verena.felder@uni-konstanz.de

Our investigation focuses on Swedish, where the tones are traditionally labeled as Accent 1 and Accent 2 (henceforth, Acc1 and Acc2). Crucially, the tonal contrast is observable only in words that minimally contain a trochee—that is, a stressed syllable followed by an unstressed syllable. Monosyllables and words with final stress are always Acc1—for example, *hund*, “dog,” *idé*, “idea.” The phonetic implementation of the accents differs from dialect to dialect. Hence, we will restrict the following description to Stockholm Swedish, which is also the language and subject background used in the experiment.

Phonologically, the tonal contrast is best described as low (Acc1) versus high (Acc2) tones, linked to the stressed syllable. Acoustically, Acc1 begins with a rise on the first syllable, whereas Acc2 has a fall (Figure 1). In addition, Acc2 has a second peak in the poststressed syllable. Thus, the difference is twofold: (1) the pitch contour in the stressed syllable and (2) the number of peaks.

There are three possible assumptions concerning the lexical specification of the tonal contrast, all of which are attested in the literature: (1) Both accents are specified (Bruce, 1977; Gussenhoven & Bruce, 1999); (2) Acc2 is the lexical accent, particularly because it is phonetically more complex (Elert, 1972; Riad, 1998, 2003), whereas Acc1 is the default; and (3) Acc1 is the lexically specified accent (Lahiri, Wetterlin, & Jönsson-Steiner, 2005), whereas Acc2 is the default.

All three approaches assume that monosyllabic words are not specified and surface with Acc1. Theory 1 (Bruce, 1977) specifies both accents and derives the correct one on grounds of information concerning the stress pattern, the lexical status, and morphological patterns of the word form in question. Theories 2 and 3 attempt to simplify this by specifying only one member of the contrast. Both of them assume that the lexical information for the speci-

fied accent is carried by word stems and morphemes. If no such information is attached, the word will surface with the respective default accent form. However, these two theories are not equally parsimonious in terms of additional assumptions needed, and, therefore, we favor the third view with the assumption that Acc1 is the lexically specified accent, which avoids setting up additional rules. We will illustrate this with the following example. The word *skriv*₁, “you write!,” which is also the verb root, is produced with Acc1 simply because it is monosyllabic. But the suffixed infinitive form, *skriv-a*₂, has Acc2. Theory 2 assumes that this is the case because the infinitive suffix *-a* is lexically specified for Acc2 and governs the word accent. However, this explanation runs into difficulties as soon as a prefix such as *be-* is added to form the derived word *be-'skriv-a*₁, “to describe.” If the suffix *-a* governed the word accent, *be-'skriv-a*₁ should have been Acc2, but instead it is Acc1. In order to arrive at the correct output form, Theory 2 needs additional rules for deleting the incorrect Acc2. By assuming that only Acc1 can be specified for affixes and words, Theory 3 avoids such problems. Again, *skriv*₁ is Acc1, not because it is specified as such, but simply due to its being monosyllabic. All unaccented words with a trochaic foot (a stressed syllable followed by an unstressed syllable) are assigned Acc2 as default. The inflected word *skriv-a*₂ takes Acc2 by default because it is not specified for accent (i.e., the suffix *-a* is unspecified) and provides the necessary trochaic structure for Acc2 to be realized. Instead, the prefix *be-* is specified for Acc1's governing word accent, and *be-'skriv-a*₁ is assigned Acc1. There is no further need for additional rules. A consequence of this analysis is that words and affixes specified for Acc1 are always Acc1, and thus anything specified in the mental lexicon will not lose this specification by a later process.

Note that Theory 3 does not specify all Acc1 words as such in the mental lexicon. As has already been mentioned, monosyllabic words are unspecified. This is also true for words that are monosyllabic only in the lexicon (i.e., underlyingly) but are produced as disyllabic words due to epenthetic vowel insertion when word final. Epenthetic insertion of a vowel means that, as a word is uttered, the speaker inserts a vowel (usually a schwa) between two adjacent consonants. In Swedish, this happens to words ending in consonant + sonorant consonant clusters (/Cl/, /Cr/, /Cn/), such as /humr/ “lobster,” which are produced with a schwa between the word-final sonorant consonant and the preceding consonant, resulting in *hummer*₁. In the mental lexicon, these words are stored without the inserted vowel—that is, in the Swedish case, as monosyllabic words (see Kristoffersen, 2000, for Norwegian, which has exactly the same phenomena; see Riad, 1998, for Swedish). However, they are always written as disyllabic words, and native speakers of Swedish are completely unaware of their monosyllabic character (<http://lexin2.nada.kth.se>;¹ Bruce & Hermans, 1999). Their monosyllabic underlying form *humr-* is visible in suffixed forms such as the plural *humr-ar*, never in isolation. There are similar cases in English, too. For instance, the English word *bottle* is considered to be a disyllabic word by native speakers (and dic-

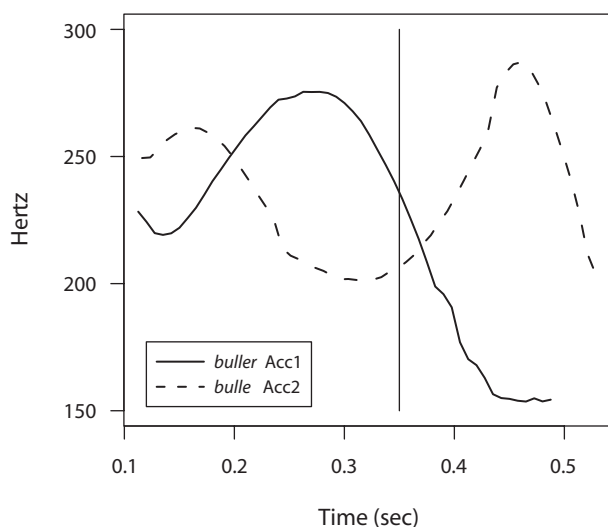


Figure 1. Pitch tracks of a pair of Accent 1 (Acc1) and Accent 2 (Acc2) words, with segmentally identical first syllables: *buller*₁ (solid line) and *bulle*₂ (dotted line). The vertical line marks the syllable boundary. The words were spoken by a native speaker of Stockholm Swedish.

tionaries). In fact it is underlyingly monosyllabic, which is seen, for instance, in the verb *bottling* with the addition of a suffix *-ing*. In contrast, *counsel* is truly disyllabic, and consequently, trisyllabic *counselling* is perfectly possible, but **botteling* is impossible. As is true for Swedish speakers, English natives would characterize both words, *bottle* and *counsel*, as disyllabic. That is, words that are monosyllabic in the mental lexicon but disyllabic on the surface are always Acc1 in the singular stem form in Swedish. However, they are not specified as Acc1 words but are assigned Acc1 by default, due to their monosyllabic structure. In combination with suffixes that turn them into disyllabic words with a trochaic structure, it can happen that they will surface with Acc2, as, for example, in the plural form *humr-ar*₂. However, a word that is lexically specified for Acc1 will remain Acc1 even if the plural suffix *-ar* is added—for example, *abbot*₁–*abbot-ar*₁ “abbot.” The only Acc1 stems in Swedish that are underspecified for accent are underlyingly monosyllabic words (with or without disyllabic structure in production) and words with main stress on the final syllable (because Acc2 needs a poststress syllable for the second peak).

In sum, we assume that stems and affixes can be lexically specified only for Acc1. The default rule is that a word with a trochaic structure receives Acc2, whereas a monosyllabic word or a word with final stress can only become Acc1. The only exceptions to this rule are trochaic words—that is, words that provide the necessary structure for Acc2 assignment, which nevertheless take Acc1. These words have to be lexically specified for Acc1. This logic of assigning the correct accent in production is outlined in Figure 2.

Perceptually, native speakers have no difficulty in differentiating between words that contrast in tone (Bruce, 1977; Fant, 1973). To our knowledge, there is only one study, by Efremova, Fintoft, and Ormestad (1963), on East Norwegian, in which the amount of information neces-

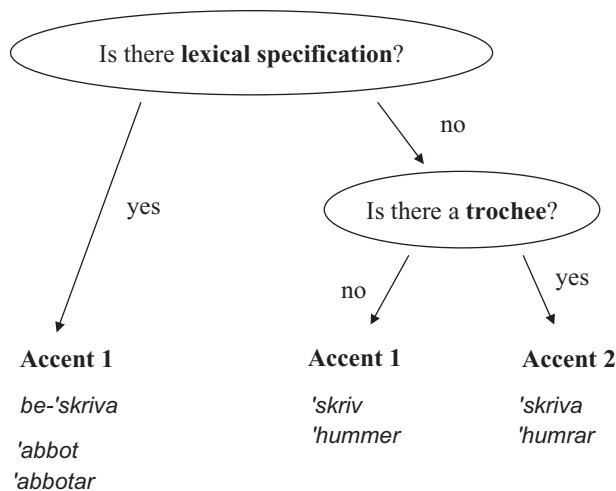


Figure 2. Diagram of word accent assignment based on Lahiri, Wetterlin, and Jönsson-Steiner (2005). The examples given in italics correspond to the cases mentioned above.

Table 1
Example of the Experimental Design

Auditory Fragment	Visual Target	
	Accent 1	Accent 2
[ham] ₁ or [ham] ₂	hambo _{specified}	hampa _{unspecified}
[hum] ₁ or [hum] ₂	hummer _{unspecified}	humla _{unspecified}

sary to identify the accents has been examined. In a gating study, they showed that subjects were already very accurate in distinguishing between Acc1 and Acc2 during the stressed vowel of the first syllable. However, there has been no experimental work on the consequences of tone specification for lexical access.

Our contention is that the surface contour should positively affect and, therefore, hasten the recognition of its lexically specified accented word but have no such speeding-up effect if the matching accent is not lexically specified. Thus, we predict that Acc1 contours should accelerate the speed of recognition of matching lexically specified Acc1 words, but lexically unspecified Acc1 words or Acc2 words will not be affected by a corresponding Acc1 or Acc2 contour, respectively. This prediction results from the assumption that perceiving the accent of a word can assist in lexical access only if the lexical entry provides any information on the tone of the word in question. Only lexically specified Acc1 words have information on their tonal pattern in their lexical entries. Unspecified Acc1 and Acc2 words cannot contribute to word processing with lexical information on tone. Hence, the search process can be terminated more quickly for specified Acc1 words than for the others. Theory 2, which specifies Acc2, would, if anything, expect Acc2 words to be recognized more quickly, and it would not differentiate Acc1 words into two groups with different behavior. For Theory 1, where both accents are assumed to be specified, there should be no asymmetry in speed of recognition.

We tested these hypotheses in a cross-modal forced choice experiment in which listeners were presented auditorily with one of the segmentally identical first syllables of a pair of words differing in accent, followed by the two visually presented words (see Table 1). Listeners then had to decide which of the two words the auditory fragment was taken from. Since we expected lexical specification of tone in the mental lexicon to govern the speed of processing, we hypothesized a faster identification for specified Acc1 words such as *hambo*₁ than for unspecified Acc1 words such as *hummer*₁ and no difference between the corresponding sets of Acc2 words (*hampa*₂, *humla*₂). Neither did we expect a difference between unspecified Acc1 words and Acc2 words.

EXPERIMENT 1

Lexical Accent Influence on Forced Choice Task

Method

Stimuli. A total of 60 pairs of familiar disyllabic Swedish nouns served as stimuli in the experiment. The members of each pair shared the segments of the first syllable but differed in accent. They

Table 2
Word Frequency and Cohort Size Means
(With Standard Deviations)

Word	Word Frequency				Cohort Size (Lexicon Count)	
	Corpus Data		Subjective Rating		<i>M</i>	<i>SD</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>		
Acc1-S	715	1,287	2.86	1.34	5.46	3.80
Acc1-U	614	863	3.23	1.29	6.32	4.94
Acc2-CS	1,190	2,879	3.14	1.34	4.58	3.97
Acc2-CU	2,311	4,438	3.51	1.31	5.93	4.40

Note—Word frequency is given separately according to corpus data and subjective ratings. Cohort size was ascertained by lexicon counts.

were divided into two sets: (1) specified Acc1 words (henceforth, Acc1-S) with their segmentally corresponding Acc2 counterparts (i.e., **hambo**₁ ~ **hampa**₂) and (2) unspecified Acc1 words (henceforth, Acc1-U) with corresponding Acc2 words (i.e., **hummer**₁ ~ **humla**₂). Although Acc2 words are always unspecified, to indicate the segmental overlap within the pairs, the words corresponding to Acc1-U and Acc1-S are indicated in the following by Acc2-CU and Acc2-CS, respectively, where the C stands for *corresponding to* and the U refers to the set of *unspecified* Acc1 words, and the S refers to the set of *specified* Acc1 words (e.g., **hampa**₂ is Acc2-CS, **humla**₂ is Acc2-CU). Thus, half of the Acc1 words used in the experiment are assumed to be lexically specified (i.e., Acc1-S, **hambo**₁), and the other half are assumed to be underlyingly monosyllabic, only pronounced as disyllabic due to vowel insertion (i.e., Acc1-U; /**humr**/ > **hummer**₁). The 60 Acc2 words were unspecified for accent.

Frequency. The choice of words was restricted by the requirement that disyllabic monomorphemic nouns should differ in accent but share the segmental structure of the first syllable. Therefore, word frequency could not be fully balanced. Word frequency was assessed using corpus data as well as subjective ratings. The online språkbanken corpus (<http://spraakbanken.gu.se>) contains more than 132 million words collected from Swedish newspapers and novels dating from 1965 to 2004. Word frequency is indicated by the total number of occurrences in the corpus, shown in Table 2. A Wilcoxon rank sum test was conducted for various group comparisons. As can be seen in Table 3, although the frequencies differed, none of the comparisons reached statistical significance.

In order to ascertain subjective frequency impressions, 10 native Swedish speakers who did not participate in the experiment were asked to indicate on a scale from 1 to 5 how often they thought the words were used in today's Swedish, with 1 meaning *never* and 5 *very often*. The words were presented to them visually in a randomly ordered list. Mean responses are shown in Table 2, statistical results in Table 3. This time, all the comparisons reached significance. Note, however, that the actual difference between their means is very small. Nevertheless, Acc1 words are less frequent than Acc2 words. Acc1-S words are less frequent than Acc1-U words, and both are less frequent than their corresponding Acc2 words.

We did not consider these results detrimental for the experiment, because our hypothesis states that Acc1-S words are responded to

Table 3
Statistical Results for Various Comparisons of
Word Frequency in Corpus Data and Subjective Ratings,
Using Wilcoxon Rank Sum Tests

Comparison	Corpus Data		Subjective Rating	
	W Test	<i>p</i> Value	W Test	<i>p</i> Value
Acc1-S vs. Acc1-U	397	.57	37,890	<.001
Acc1-S vs. Acc2-CS	340	.15	39,815	<.02
Acc1-U vs. Acc2-CU	341	.22	39,425	<.01
Acc2-CS vs. Acc2-CU	385	.58	37,799	<.001

most quickly. Usually a high word frequency leads to lexical decisions faster than those for low-frequency words. Acc1-S words have the lowest frequency, and consequently, frequency would not explain the predicted faster responses to the very same set of items.

Cohort size. Another factor that could theoretically influence speed of recognition is cohort size. Cohort size was ascertained with a lexicon on Swedish pronunciation (Hedelin, 1997). We counted all words with initial stress starting with the same segments as the nouns in the experiment. Syllable structure and word class were not taken into account, whereas accent, vowel length, and quality were. Compounds and derived words were not counted. Mean cohort size can be seen in Table 2. Again, we assessed potential differences with Wilcoxon rank sum tests. Cohort sizes did not differ between specified and unspecified Acc1 words, between their corresponding sets of Acc2 words, or between Acc1 and Acc2 words. Also, when accent was not considered in counting the cohort, there was no difference between the specified and the unspecified sets. All *p* values were above .13.

Stimulus preparation. The first syllables of the experimental words served as auditory prime fragments. All the words were recorded in Stockholm, read by a female native Stockholm speaker. The words of interest appeared in focused position at the end of the sentence *Nej, han sa' inte bulle, han sa' buller* (No, he did not say bun, he said noise). Recordings were done with Cool Edit (Syntrilium Software Corporation, Phoenix, AZ), offline editing in Praat (Boersma & Weenink, 2006). The first syllable was cut out of each sentence-final word at zero crossings before transitions to the segments of the second syllable. Both first syllables of each stimulus pair had exactly the same duration in milliseconds. This was not achieved by manipulating the recorded speech signal in any way, but simply by a careful choice of the position of the cut. This implies that fragments were of the same duration as a whole, and we did not adjust the lengths of the vowels and consonants that made up the fragments. The amount of *f*₀ information available to the listeners (between an average of 300 msec of pitch contour in Acc2-CU fragments and 314 msec of average pitch contour in Acc2-CS fragments) was statistically the same for all the conditions. Thus 120 first syllables of originally disyllabic words resulted; half of them carried Acc1, and half Acc2 information. These 120 syllables formed 60 pairs with the same durational and segmental information each, the only difference being found in *f*₀ contour. Great care was taken to avoid coarticulatory information in the prime fragments that could bias subjects toward one of the two words irrespective of accent information. In order to ensure that the Acc1-S and Acc1-U fragments were not fundamentally different with respect to their tonal contours, we calculated the *f*₀ values of the beginning, middle, and end of the vowel of each fragment. Within the Acc1 sets, the shape of the *f*₀ contour—assessed as a ratio of the midpoint and beginning, as well as of the end and the midpoint, of the vowel—was exactly the same for the Acc1-S and Acc1-U fragments ($F = 0.01, p = .93$). The duration of the *f*₀ contours across all pairs did not differ ($F = 0.17, p = .92$).

Experimental design. Nine participants could be tested simultaneously in a quiet room. Instructions were given orally and in written form. For each trial, the participants first heard a short beep via headphones in order to focus their attention. They heard a word fragment 200 msec later, and immediately at its offset, two words were projected (Liesegang dv245) next to each other (approximately 1 m apart) on the wall and remained there for 4 sec. Both words' first syllables had the same segments as the auditory fragment, one of them being Acc1, the other Acc2. For instance, if the fragment "ham" (which could carry either Acc1 or Acc2) was presented auditorily, the two words **hambo** and **hampa** were projected (see Table 1). The participants' task was to identify the word the syllable matched with and press the left or right button of a response box, depending on which side the target word was placed on the screen. The reaction time measurement started at the onset of the visual stimuli on the screen. The placement of the matching word was balanced for left and right, controlled for accent and stimulus pairs. Each fragment

was presented once; that is, each subject was exposed to a total of 120 trials. Consequently, each word pair appeared twice, once with the Acc1 fragment and once with the Acc2 fragment. The exact order of presentation was randomized. The whole experiment lasted approximately 10 min.

Subjects. Twenty-nine students from Stockholm University, all native speakers of standard Swedish (mean age, 31 years), who were brought up in or around Stockholm, consented to be subjects and were paid for their participation.

Results

One of the 29 subjects was excluded because the majority of his responses were outside the time window of 3 sec. Two pairs of words were removed because one member was considered to have an optional accent specification by some native speakers. An additional four pairs were excluded from the analyses because one member of each pair yielded chance responses over all subjects. Responses below 200 msec were excluded, since these responses could only have been guesses (in all 14 responses). Only those items of which both members of the pair were correctly responded to were considered for the reaction time analysis.

Although pairs of Acc1 and Acc2 words were always matched for initial segments, there were unavoidable segmental differences between words belonging to the specified and unspecified sets (see Table 1). To control for any possible segmental differences across the two sets, Acc2 words remained split into Acc2-CS and Acc2-CU throughout all the analyses.

Response accuracy. Due to a slightly different number of trials within each condition, we ran an ANOVA with a restricted maximum likelihood (REML) analysis (the type of ANOVA recommended in such cases) in JMP (SAS, Cary, NC), with the arcsine transformed ratio of correct versus incorrect responses as a dependent variable and accent (Acc1 vs. Acc2) and specification set (-S and -CS vs. -U and -CU) as independent variables. This resulted in a significant main effect for accent [$F_1(1,27) = 10.95, p < .01$].² Acc1 words were responded to more accurately than Acc2 words. Response accuracy was not affected by specification (see Figure 3).

Response speed. A further REML analysis was run with reaction time as a dependent variable and accent (Acc1 vs. Acc2) and specification set (-S and -CS vs. -U and -CU) as independent variables. This yielded a significant main effect for subjects and a marginally significant main effect for items for both accent [$F_1(1,27) = 11.83, p < .01; F_2(1,104) = 3.74, p = .06$] and specification set [$F_1(1,27) = 13.10, p < .01; F_2(3,104) = 3.06, p = .08$], as well as a significant accent \times specification set interaction for subjects [$F_1(1,27) = 7.44, p < .02; F_2(1,104) = 1.59, p = .21$]. Holm-corrected t tests revealed that Acc1-S words were responded to significantly more quickly than was any other accent type. Responses were faster for them than for Acc1-U words [$t(1,27) = -4.32, p < .001$] and also for Acc2-CS [$t(1,27) = -4.10, p < .001$] and Acc2-CU [$t(1,27) = -4.99, p < .0001$] words. Although Acc1-S words differed significantly from their Acc2-CS counterparts, Acc1-U words did not differ from Acc2-CU words [$t(1,27) = -0.76, p = .45$]. Fur-

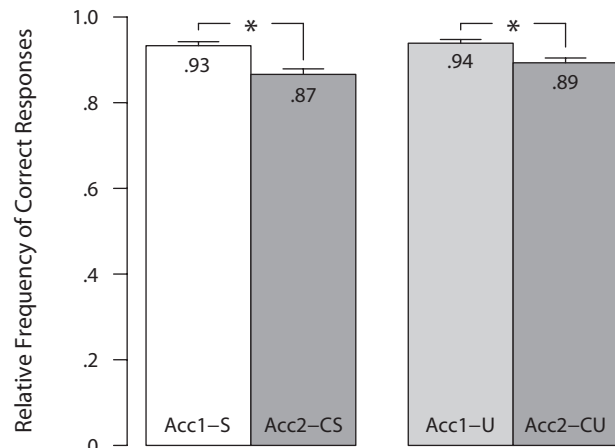


Figure 3. Proportion of correct responses. The columns indicate the four groups of items. Asterisks indicate significant differences ($p < .05$). The error bars above each column point to standard errors.

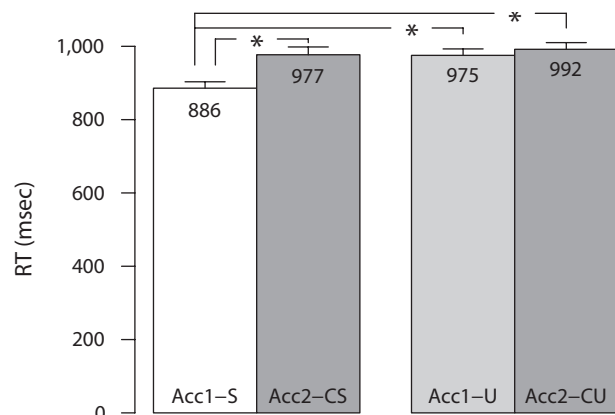


Figure 4. Reaction time (RT) results in milliseconds. The columns indicate the four groups of items. Asterisks indicate significant differences ($p < .05$). The error bars above each column point to standard errors.

thermore, Acc2-CS and Acc2-CU words did not differ [$t(1,27) = -0.82, p = .42$; see Figure 4].

Discussion

In the forced choice experiment above, we tested the impact of lexical specification of tone on speed and accuracy in word identification. The subjects heard the first syllable of a word (e.g., *ham*₁- from *hambo*₁) and then were presented with two words on a screen (in this case, *hambo* and *hampa*) and were asked to decide which of the words the auditory fragment was taken from. The accent information for the auditory fragment was intended to be the only cue for the subjects' decisions. Acc1 and Acc2 words were identified with high precision, confirming that the tonal information of the first syllable is sufficient for listeners to correctly identify the accent. Interestingly enough, Acc1-S words were identified considerably more

quickly than Acc1–U words (despite having the same tonal contour) and also more quickly than Acc2 words (despite having the same segmental information), whereas there was no difference in reaction times between the latter two. In other words, those items that we assumed to be specified for Acc1 in the mental lexicon were identified more quickly than items that were unspecified, irrespective of their accent information. For further discussion of these results, see the General Discussion section below.

In order to draw the conclusion that these results were caused solely by tonal information and its mental representation, we had to make sure that there was no response-biasing segmental information available in the signal. Recall that both members of a pair had the same segmental information in their first syllable (e.g., *ham* from *hambo*, *hampa*). However, segmental information in the second syllable was different between the two members of each pair. Therefore, the possibility of coarticulatory information in the first syllable, which could give an inkling of the segments in the second syllable, had to be ruled out. We therefore conducted a separate forced choice experiment.

EXPERIMENT 2

Validating the Accent Specification Hypothesis: Segmental Coarticulation or Tonal Influence?

Method

Stimulus preparation. In order to assess the impact of segmental cues on word choice, we presented subjects with incrementally increasing stimuli (similar to gating) in a way similar to that in the forced choice task itself. Since we were interested in a possible segmental bias, we had to eliminate the impact of tonal information. To achieve this, first and foremost, the tonal information was removed from all the stimuli. Our rationale was that if segmental coarticulatory information biased subjects toward one of the two response alternatives, independently of tonal information, this bias should be reflected in their responses to stimuli that lacked the tonal contour. If, for instance, subjects responded more quickly to Acc1–S words due to more coarticulatory information in these items, this bias should become apparent when they were presented with stimuli without a tonal contour.

Using Praat, the original f_0 was deleted and replaced by a constant 200-Hz pitch line. Next, stimulus extracts of four different lengths were cut out from each word pair. The shortest extract consisted of the initial consonant and half of the vowel—that is, up to half of the vowel of a CV sequence such as *mu-* from *muskel*. The second extract size was identical to the fragments presented in the main experiment (in fact, the very same auditory files were used, only with flattened pitch) and thus consisted of either a CV (with full vowel) or a CVC sequence (e.g., *mus-*), depending on syllable structure. The third extract included the first consonant of the second syllable [i.e., CV(C)C; e.g., *musk-*], and the fourth went up to the middle of the vowel in the second syllable [i.e., CV(C)CV; e.g., *muske-*]. Thus, for each word, the length of the extract increased with each step. The crucial set of stimuli were those with the same length as the fragments in the original experiment, since we were interested in whether these first-syllable fragments contained subsegmental cues as to the second syllable. The shorter extracts were included to ensure that the stimuli were indeed ambiguous at the onset of each word. The longer two extracts were included for reasons of comparison, since with increasing fragment length, more and more stimuli can be differentiated from their corresponding member.

Experimental design. The subjects were presented with an auditory word extract and were asked to decide which of two words it be-

longed to (e.g., *ham*–*hambo*, *hampa*). The two target words appeared 1,000 msec before the auditory word extract and stayed on the computer screen until a response was made. The pitch-flattened extracts were presented to them in four blocks of 120 trials each, with all extracts of the same length in one block. The extracts were randomized for each block separately. The block with the shortest extracts was presented first, the one with the longest last. Half of the correct responses appeared on the left side of the screen, the other half on the right side. For each word pair, the position of the words on the screen stayed constant within a block and was balanced across blocks.

The reasons for the change in the presentation mode of the first experiment were twofold. Our emphasis was to observe (1) whether lack of tonal information on the fragments presented in Experiment 1 would have an adverse effect on the correct recognition of the words, and (2) how much acoustic information would be necessary for the subjects to make the choice between the words. We deliberately chose to allow the subjects to monitor the speed of the experiment, since as the extracts increased in length, more segmental material was available and the decisions would become easier. However, the mode of presentation was changed (visual words presented before the auditory stimuli) because, in a pilot study, the subjects reported that they had tremendous difficulty, with the short extracts, keeping the exact acoustic form in mind without knowing what alternatives they would be given. This led them to press the buttons randomly (these subjects were not included in the statistics). Consequently, we decided to make the task easier by providing the two word alternatives before playing the acoustic extract and allowing the subjects enough time to read both alternatives.

We hypothesized that the subjects would not be able to decide which word the extracts belonged to for the shortest fragments, since they were cut well before the syllable boundary. With sufficient segmental information in long extracts, they should be able to identify most of the words. The crucial condition consisted of the fragments with the same length as in the main experiment. We expected that these fragments should not allow the subjects to predict the correct word on the basis of segments. Or, if they sometimes did respond correctly, all accent types should be affected in a similar fashion. In the main experiment, Acc1–S words were identified more quickly than Acc1–U words. If this was due not to the specification of tone but, rather, to segmental information biases, Acc1–S words should contain more segmentally biasing information than should Acc1–U words, and consequently, Acc1–S words should be correctly identified more often than Acc1–U words on the basis of segmental information only. Therefore, if, with flattened pitch contour, Acc1–S words are identified more correctly than Acc1–U words, this speaks in favor of a segmental, rather than a tonal-representational, explanation of the results in the main experiment.

Subjects. For the gating experiment, 12 native speakers of Swedish living in Zurich were recruited and were paid for participation. The subjects had left Sweden in their late 20s and continued to use their native language regularly at home. They were informed that any pitch information that would naturally occur on items that they were going to hear had been removed. Hence, they were asked to concentrate solely on segmental information and decide, as accurately as possible, from which of the two words on the screen the auditory fragment was taken. Since tonal information was removed from the stimuli, it was no longer crucial to have speakers only from Stockholm. In the main experiment, this was important, due to tonal variation across dialects.

Results

The percentages of correct responses are given in Table 4.

An ANOVA was conducted for each block in JMP (SAS, Cary, NC), with the arcsine transformed ratio of correct responses as a dependent variable and accent (Acc1 vs. Acc2) and specification set (–S and –CS vs. –U and –CU) as independent variables. Subjects was included as a ran-

Table 4
Percentages of Correct Responses for the
Four Different Sizes of Extracts

Fragments	CV-	CV(C)-	CV(C)C-	CV(C)CV-
All	49	53	71	93
Acc1-S	54	70	75	98
Acc1-U	51	75	79	88
Acc2-CS	43	35	68	98
Acc2-CU	49	32	63	90

Note—Statistics are given for all extracts together, as well as separated for the four types of accent words used. Results for the extracts with the same segmental information as the fragments in the main experiment are in bold.

dom variable. In the first block with the shortest extracts, no significant effects were obtained. In the second block, with the extracts of segmental information identical to the fragments in the main experiment, there was a main effect for accent [$F_1(1,11) = 11.36, p < .01$], due to more accurate responses to Acc1 words, as well as an accent \times specification set interaction [$F_1(1,11) = 6.70, p < .05$]. Post hoc t tests revealed a slight difference between Acc1-S and Acc1-U words [$t(1,11) = -2.58, p < .05$]. Acc1-U words were classified correctly in 75% of the cases, Acc1-S words in 70%. The analysis of the extracts of the third block (with segmental information of the onset consonant of the second syllable) revealed a main effect for accent only [$F_1(1,11) = 7.65, p < .02$], again due to more accurate responses to Acc1 words. Finally, analysis of the longest extracts led to a main effect of specification set [$F_1(1,11) = 89.04, p < .0001$], because of more accurate judgments for the specified set, independent of word accent.

Discussion

We conducted a second forced choice experiment similar to the first experiment in order to check whether the speeded responses to Acc1-S words were, indeed, due to lexical specification of Acc1 or to an accidental advantage of these fragments due to coarticulatory segmental information. Hence, the tonal contour was removed from the stimulus material, and extracts of increasing length were presented to the subjects. The rationale was that if segmental information led to the results, Acc1-S words should be identified more accurately than Acc1-U words on the basis of segmental information only.

The results showed a main effect for accent for the two extracts of intermediate lengths. The subjects in the experiment reported that a flattened pitch contour resembled Acc1 more than Acc2. Therefore, they more often chose the Acc1 word in cases in which segments did not give sufficient information. This perception bias was confirmed by native Swedish speakers who were familiar with the theory of tone in Scandinavian. Consequently, this main effect does not tell us that there was a segmental bias toward Acc1 words in the main experiment but that the bias was due to a response bias toward Acc1 words in cases of flat pitch contour. This implies that we cannot compare Acc1 with Acc2 words but can still compare Acc1-S with Acc1-U words. In this comparison, Acc1-S words should not be identified more accurately than Acc1-U words on the basis of segmental information only. Otherwise, faster

responses to Acc1-S words in the main experiment could be explained by a segmental advantage of these Acc1-S targets, as compared with Acc1-U targets.

The accent \times specification interaction for the fragments that were also used in the main experiment was due to a difference of 5% in accuracy between Acc1-U and Acc1-S words, where responses to the latter were less accurate than those to the former. Thus, Acc1-U words with flattened pitch were recognized more correctly than Acc1-S words with flattened pitch. Note that these accuracy results are, in a sense, the opposite of the results in the main forced choice experiment, in which responses to Acc1-S words were the fastest. If anything, this result tells us that any segmental cue in the fragments is more prominent in Acc1-U words than in Acc1-S words. The faster responses for Acc1-S words, as compared with Acc1-U words, in the main forced choice experiment cannot be attributed to effects of coarticulation or any other segmental cue. Otherwise, a bias in the segmental makeup of these toneless fragments would have affected the correctness of the responses, as was observed for the third and fourth extract lengths.

The main effect of specification for the longest extracts is due to the fact that the members of a pair deviate from each other in terms of segmental information at different points in time. In the specified set, 19 word pairs deviate segmentally at the onset consonant of the second syllable, 9 deviate at the vowel of the second syllable, and 1 pair differs only in the final consonant. In the unspecified set, 19 word pairs deviate segmentally at the onset consonant of the second syllable, 3 deviate at the vowel, and 7 differ only in the last consonant. Since the longest extract lasted up to the vowel of the second syllable, only 1 word pair in the specified set but 7 word pairs in the unspecified set stayed ambiguous in terms of segments. This explains the less correct identification of the longest extract in the unspecified set with a flattened pitch contour.

We have not reported reaction time data for several reasons. First, whereas in the main experiment, the subjects were presented with the auditory stimulus first, followed by a visual presentation of the word pair, this design proved too difficult in the control study, as outlined above. Hence, we presented the target words 1,000 msec before the onset of the acoustic fragment. Furthermore, the subjects were not instructed to respond as quickly as possible, since our main interest was in the accuracy of their responses. Also, the words stayed on the screen until the response was made, so there was no time pressure as such. This led the subjects to respond very slowly and with a high variability in response speed. Reaction time data would not be indicative and, hence, was not analyzed.

GENERAL DISCUSSION

The central issue raised here is whether the mental representation of tone governs the speed and accuracy with which listeners identify word accents. Presented with segmentally identical first syllables differing in accents, Swedish listeners had to decide which of the two visually presented disyllabic Acc1 and Acc2 words the auditory fragment was

part of. The high accuracy rate of the responses (overall 91% correct) indicates that the pitch contour of the initial syllable was salient. Recall that the Acc1 words were of two types, specified disyllabic words (Acc1-S) and monosyllabic stems, which became disyllabic by vowel insertion and thus were assumed to be unspecified (Acc1-U). Acc2 words, considered to be lexically unspecified for tone, were segmentally matched for initial syllables with their Acc1 counterparts—Acc2-CS, Acc2-CU.

We predicted that on hearing a surface Acc1 tonal contour, the recognition of matching Acc1 words would be accelerated, provided that they were lexically specified. At the same time, the lexically unspecified Acc1 and Acc2 words could not benefit from the tonal information equally well. This prediction was based on the assumption that perceiving the accent of a word can assist in lexical access only if the lexical entry provides information on the tone of the word in question and that only lexically specified Acc1 words have information on their tonal pattern in their lexical entries. Or then, the lexical decision process is faster for words that have accent information stored directly in their lexical entry than for those words for which their accent identity has to be determined by nonlexical rules. Hence, in the main experiment, we predicted faster responses to matching Acc1-S words than to Acc1-U or Acc2 words and no difference between the unspecified forms.

Accuracy results in the main experiment showed that all the words were recognized very accurately, Acc1 words even more so than Acc2 words. There was no difference in accuracy within these two groups. We assume that Acc2 words were recognized less correctly because the peak of the second syllable was missing, leading to a quite unusual and incomplete percept of Acc2, whereas most information for Acc1 was provided in the auditory fragment.

As was predicted, responses to Acc1-S words were faster than those to Acc1-U, Acc2-CS, and Acc2-CU words. Furthermore, there was no difference in response speed between the latter three. Faster responses to Acc1-S words are in favor of our assumption that specification in the lexicon speeds lexical access and word recognition. The fact that there is no difference between Acc1-U and Acc2-CU words supports our claim that these two unspecified classes are similar with respect to the specification of lexical tone, in spite of the fact that the acoustic tonal patterns are very different.

Differences in word frequency cannot account for these results. Words with high frequency are usually reacted to more quickly than words with low frequency (Taft, 1979). We observed the shortest reaction times to Acc1-S words, which were subjectively rated as least frequent. The Acc1-U and Acc2 words differed in the subjective frequency ratings but elicited equally "slow" responses.

Furthermore, the results cannot be explained by a speed-accuracy trade-off. Acc1-S words not only had faster responses than did Acc2-CS words; they were also identified more correctly. In addition, there was no difference in accuracy between Acc1-S and Acc1-U words, and still Acc1-S words were responded to considerably more quickly.

One could argue that the lack of the second tonal peak in the auditory Acc2 fragment not only caused the subjects to respond less correctly, but also slowed down their response speed. This would serve as an alternative explanation for the difference in speed between Acc1-S and Acc2-CS words. However, it is no explanation for the difference in reaction time between Acc1-S and Acc1-U words, nor can it explain why responses to Acc1-U words are equally slow as those to Acc2-CU words.

Furthermore, the results cannot be explained by a one-to-one matching of the physical shape of the pitch contour, since the surface contours of the Acc1 words, specified and unspecified, were the same. Nevertheless, the Acc1-S words were significantly faster to access when the matching contour was heard than were the Acc1-U words. Despite having the same f_0 contour, there was an inherent difference between the Acc1-S and Acc1-U words in terms of their lexical representation: The Acc1-U words were underlyingly monosyllabic. Could this have slowed down their recognition? Had syllable structure been an issue, we would have expected there to be a difference between monosyllabic Acc1-U words and disyllabic Acc2-CU words, and there was none. Furthermore, responses to specified Acc1-S words were significantly faster than those to Acc2-CS words, although they were both underlyingly disyllabic. One possibility is that a combination of slowing effects, caused by the lack of the second tonal peak for Acc2 words and the monosyllabicity of Acc1-U words, led to the present results. Additional work has to be done to rule this out.

The second experiment rules out segmental cues as a reason for faster identification of Acc1-S words. The differences in reaction time could have been caused by segmental information only if the fragments of words specified for Acc1 had accidentally more coarticulatory cues than did fragments of either unspecified Acc1 or Acc2 words. In this case, Acc1-S words would have been identified more quickly, because the acoustic information gave an additional hint or even served to exclude the alternative Acc2 word in the forced choice task. The results of the gating study indicate that Acc1-S words were cued even less by segmental information than were Acc1-U words. The difference between the two was rather small and, if anything, accelerated responses to Acc1-U words, rather than to Acc1-S words.

Nor can these results be explained as an artifact of orthography: For historical and structural reasons, there are certain correspondences between final segments in disyllabic words and their accents. For instance, disyllabic words ending in *-a* are usually Acc2, whereas those ending in *-i* are usually Acc1-S. However, the only really homogeneous group in terms of final segments are the Acc1-U words, because the final consonant of these words is either /l/, /r/, or /n/, since underlyingly they all end in C+sonorant clusters. Nevertheless, such clusters can also be part of Acc2 words and Acc1-S words. In any event, had orthography had an effect, it would have speeded up responses to Acc1-U words, due to their homogeneous endings. Acc1-S words do not stand out orthographically.

Thus, it appears that none of the alternative explanations attempting to account for the results hold. It seems that our assumptions concerning the effect of lexical specification of Acc1–S words on processing are borne out. The fact that the same Acc1 contour leads to a faster decision for Acc1–S words than for Acc1–U words suggests that their abstract lexical specification facilitates their processing, thereby making it easier for subjects to come to a decision. Theory 2 (Elert, 1972; Riad, 1998, 2003), which assumes Acc2 words to be marked, cannot explain the difference between Acc1–S and Acc1–U words on grounds of their underlying specification.

Lahiri, Wetterlin, and Jönsson-Steiner (2005, 2006; Wetterlin, 2007) assume the same logic in lexical specification also for Norwegian, again specifying Acc1. As for Swedish, other theories assume either that both accents need to be marked in the lexicon (Withgott & Halvorsen, 1984) or that Acc2 is lexically marked, whereas Acc1 is the default (Haugen, 1967; Kristoffersen, 2000). However, although Kristoffersen (2000) assumed that only Acc2 is lexically marked, due to its higher complexity, he revised this view in later work (Kristoffersen, 2006a, 2006b). In his analyses of the Norwegian dialects of North Gudbrandsdal and Oppdal, he concluded that in these dialects, Acc1 and Acc2 have the same HLH tone contour, differing only in timing. Whereas in Acc2 this contour was aligned with the word in a simple left-to-right manner, in Acc1 the L tone of the contour was aligned with the stressed syllable. He therefore concluded that Acc1 is lexically marked in these dialects, whereas Acc2 is assigned by default.

A second, independent line of evidence for the default character of Acc2 comes from studies on language acquisition. The most common argument for marking Acc2 is the complexity of this accent, having two, rather than one, high peaks. Since it is more complex, one could assume that it is also more difficult for children to acquire as they learn to talk. Hence, one would predict that children overgeneralize the less complex and default Acc1. Contrary to this, if we assume that Acc1 is lexically specified, and Acc2 is the default for any trochaic structure not otherwise specified, we further assume that children overgeneralize the default Acc2 in the process of language acquisition. That is, we would predict that if a child makes mistakes in accent assignment, it wrongly assigns default Acc2 to Acc1 words, although Acc2 is more difficult to produce. We would further assume that, later in the development, this happens more often to unspecified Acc1 words than to specified ones. These hypotheses have been confirmed by studies on Scandinavian language acquisition (Plunkett & Strömquist, 1992). At about 17 months of age, children start to produce an Acc2 contour (Engstrand, Williams, & Strömquist, 1991). Furthermore, Hellquist and Olsson (1981) found that children tend to overgeneralize Acc2, whereas there is no such trend for Acc1. Children often use both accents for the same word in consecutive trials, and there is a lot of variation between children and age groups in the extent of the overgeneralization of Acc2. Anderson (1990) reported a single case study of a boy who overgeneralized Acc2 to all disyllabic words with a post-stress syllable at the age of 22 months. None of these au-

thors had a reason to separate Acc1–S and Acc1–U words. Therefore, it would require separate analyses in order to test our second prediction that, above all, Acc1–U words will be assigned Acc2.

In the present article, we have examined for the first time the relevance of the lexical tone specification in the mental lexicon in speech perception. Our study (1) supports the notion of a nonisomorphic relationship between the speech input and the long-term representations in the mental lexicon for suprasegmental units such as tone; (2) confirms the assumption, contrary to alternative models, that only certain surface Acc1 words are lexically specified for accent and Acc2 is unspecified in the mental lexicon; and thus, (3) confirms that lexical tonal specification governs word accent identification.

AUTHOR NOTE

The research was funded by the German Research Foundation—in particular, the SFB 471 and the Leibniz Prize awarded to A.L. We are most grateful to the Department of Linguistics at the University of Stockholm—in particular, Francisco Lacerda and Thomas Riad—for their support and for providing laboratory space for the experiment. Correspondence concerning this article should be addressed to V. Felder, Department of Linguistics, University of Konstanz, 78457 Konstanz, Germany (e-mail: verena.felder@uni-konstanz.de).

REFERENCES

- ANDERSON, L. (1990). *Ett barns väg till Accent 1* [A child's way to Accent 1]. Göteborg: University of Göteborg, Department of Linguistics.
- BOERSMA, P., & WEENINK, D. (2006). *Praat*. Available at www.praat.org.
- BRUCE, G. (1977). *Swedish word accents in sentence perspective*. Lund: Gleerup.
- BRUCE, G., & HERMANS, B. (1999). Word tone in Germanic languages. In H. van der Hulst (Ed.), *Word prosodic systems in the languages of Europe* (pp. 605–658). Berlin: Mouton.
- CONNELL, B. (2000). The perception of lexical tone in Mambila. *Language & Speech*, *43*, 163–182.
- COOPER, N., CUTLER, A., & WALES, R. (2002). Constraints of lexical stress on lexical access in English: Evidence from native and non-native listeners. *Language & Speech*, *45*, 207–228.
- CUTLER, A., & CHEN, H.-C. (1997). Lexical tone in Cantonese spoken-word processing. *Perception & Psychophysics*, *59*, 165–179.
- CUTLER, A., & OTAKE, T. (1999). Pitch accent in spoken-word recognition in Japanese. *Journal of the Acoustical Society of America*, *105*, 1877–1888.
- CUTLER, A., & VAN DONSELAAR, W. (2001). Voornaam is not (really) a homophone: Lexical prosody and lexical access in Dutch. *Language & Speech*, *44*, 171–195.
- EFREMOVA, I. B., FINTOFT, K., & ORMESTAD, H. (1963). Intelligibility of tonic accents. *Phonetica*, *10*, 203–212.
- ELERT, C.-C. (1972). Tonality in Swedish: Rules and a list of minimal pairs. In E. S. Firschow, K. Grimstad, N. Hasselmo, & W. A. O'Neil (Eds.), *Studies for Einar Haugen* (pp. 151–173). The Hague: Mouton de Gruyter.
- ELERT, C.-C. (1981). *Ljud och ord i svenskan 2*. Stockholm: Almqvist & Wiksell.
- ENGSTRAND, O., WILLIAMS, K., & STRÖMQVIST, S. (1991). Acquisition of the Swedish tonal word accent contrast. In *Actes du XIIIème Congrès International des Sciences Phonétiques* (pp. 324–327). Aix-Marseille: University of Provence.
- FANT, G. (1973). *Speech sounds and features*. Cambridge, MA: MIT Press.
- GANDOUR, J. T. (2006). Tone: Neurophonetics. In K. Brown (Ed.), *Encyclopaedia of language and linguistics* (2nd ed., Vol. 12, pp. 751–760). Oxford: Elsevier.
- GASKELL, M. G., & MARSLÉN-WILSON, W. D. (2001). Lexical ambiguity resolution and spoken word recognition: Bridging the gap. *Journal of Memory & Language*, *44*, 325–349.
- GOW, D. W., JR. (2002). Does English coronal place assimilation cre-

- ate lexical ambiguity? *Journal of Experimental Psychology: Human Perception & Performance*, **28**, 163-179.
- GUSSENHOVEN, C., & BRUCE, G. (1999). Word prosody and intonation. In H. van der Hulst (Ed.), *Word prosodic systems in the languages of Europe* (pp. 233-271). Berlin: Mouton.
- HAUGEN, E. (1967). On the rules of Norwegian tonality. *Language*, **43**, 185-202.
- HEDELIN, P. (1997). *Norstedts svenska uttalslexikon*. Stockholm: Norstedts.
- HELLQUIST, B., & OLSSON, B. (1981). Grav och akut accent I barnspak [Grave and acute accent in child language]. In G. Bruce (Ed.), *Tema prosodi* (Praktisk lingvistik 6, pp. 6-19). Lund: University of Lund, Department of Linguistics.
- HYMAN, L. (2000). Tone systems. In M. Haspelmath, E. König, W. Oesterreicher, & W. Raible (Eds.), *Language typology and language universals: An international handbook* (pp. 1367-1379). Berlin: de Gruyter.
- KRISTOFFERSEN, G. (2000). *The phonology of Norwegian*. Oxford: Oxford University Press.
- KRISTOFFERSEN, G. (2006a). Markedness in urban East Norwegian tonal accent. *Nordic Journal of Linguistics*, **29**, 95-135.
- KRISTOFFERSEN, G. (2006b). Tonal melodies and tonal alignment in East Norwegian. In G. Bruce & M. Horne (Eds.), *Nordic prosody IX* (pp. 157-166). Frankfurt am Main: Lang.
- LAHIRI, A., & MARSLÉN-WILSON, W. D. (1991). The mental representation of lexical form: A phonological approach to the recognition lexicon. *Cognition*, **38**, 245-294.
- LAHIRI, A., & REETZ, H. (2002). Underspecified recognition. In C. Gussenhoven & N. Warner (Eds.), *Laboratory phonology 7* (pp. 637-675). Berlin: Mouton.
- LAHIRI, A., WETTERLIN, A., & JÖNSSON-STEINER, E. (2005). Lexical specification of tone in North Germanic. *Nordic Journal of Linguistics*, **28**, 61-96.
- LAHIRI, A., WETTERLIN, A., & JÖNSSON-STEINER, E. (2006). Scandinavian lexical tone: Prefixes and compounds. In G. Bruce & M. Horne (Eds.), *Nordic prosody IX* (pp. 167-173). Frankfurt am Main: Lang.
- PLUNKETT, K., & STRÖMQVIST, S. (1992). The acquisition of Scandinavian languages. In D. Slobin (Ed.), *The crosslinguistic study of language acquisition* (Vol. 3, pp. 457-556). Hillsdale, NJ: Erlbaum.
- RIAD, T. (1998). The origin of Scandinavian tone accents. *Diachronica*, **15**, 63-98.
- RIAD, T. (2003). Diachrony of the Scandinavian accent typology. In H. van der Hulst, J. Koster, & H. van Riemsdijk (Series Eds.) & P. Fikkert & H. Jacobs (Vol. Eds.), *Studies in generative grammar: Vol. 58. Development in prosodic systems* (pp. 91-144). Berlin: Mouton.
- SCHIRMER, A., TANG, S.-L., PENNEY, T. B., GUNTER, T. C., & CHEN, H.-C. (2005). Brain responses to segmentally and tonally induced semantic violations in Cantonese. *Journal of Cognitive Neuroscience*, **17**, 1-12.
- SEKIGUCHI, T., & NAKAJIMA, Y. (1999). The use of lexical prosody for lexical access of the Japanese language. *Journal of Psycholinguistic Research*, **28**, 439-454.
- SOTO-FARACO, S., SEBASTIÁN-GALLÉS, N., & CUTLER, A. (2001). Segmental and suprasegmental mismatch in lexical access. *Journal of Memory & Language*, **45**, 412-432.
- TAFT, M. (1979). Recognition of affixed words and the word frequency effect. *Memory & Cognition*, **7**, 263-272.
- VAN DONSELAAR, W., KOSTER, M., & CUTLER, A. (2005). Exploring the role of lexical stress in lexical recognition. *Quarterly Journal of Experimental Psychology*, **58A**, 251-273.
- WETTERLIN, A. (2007). *The lexical specification of Norwegian tonal word accents*. Unpublished doctoral dissertation, University of Konstanz.
- WHEELDON, L., & WAKSLER, R. (2004). Phonological underspecification and mapping mechanisms in the speech recognition lexicon. *Brain & Language*, **90**, 401-412.
- WINER, B. J., BROWN, D. R., & MICHELS, K. M. (1991). *Statistical principles in experimental design* (3rd ed.). Boston: McGraw-Hill.
- WITHGOTT, M., & HALVORSEN, P. (1984). *Morphological constraints on Scandinavian tone accent* (Rep. CSLI-84-11). Stanford, CA: Stanford University, Center for the Study of Language and Information.
- YE, Y., & CONNINE, C. M. (1999). Processing spoken Chinese: The role of tone information. *Language & Cognitive Processes*, **14**, 609-630.
- YIP, M. (2002). *Tone*. Cambridge: Cambridge University Press.

NOTES

1. The Swedish dictionary at <http://lexin2.nada.kth.se> transcribes the Acc1-U words as disyllabic words. Some examples: *hummer* [h'um:er], *feber* [f'e:ber], *finger* [f'ing:er], *fabel* [f'a:bel], *kakel* [k'a:kel]. The same transcription can be found for Acc2 words ending in -l, -r, or -n—as, for example, *cirkel* [²s'ir:kel]. This confirms that the listeners perceive these underlyingly monosyllabic words as disyllabic words.
2. The arcsine transformation was necessary in order to gain a normal distribution of the data, a prerequisite for any ANOVA, which is not given in plain correctness data. Since we had to transform the correct responses to relative frequencies for each subject, the item information was lost, and we could not run an item analysis (Winer, Brown, & Michels, 1991).

APPENDIX
List of Experimental Items

Acc1-S	Acc2-CS	Acc1-U	Acc2-CU
abbot	abba	buller	bulle
asbest	asfalt	ceder	zebra
cancer	kanna	cider	siden
cirkus	cirkel	fabel	fader
city	cittra	feber	fela
curry	kurva	fiber	fika
eko	eka	finger	finka
fikus	fikon	fläder	fläta
frukost	fruktan	flipper	flicka
guru	gute	galler	galge
hallick	hallon	hummer	humla
hambo	hampa	kakel	kaka
judo	jude	klotter	klocka
kassler	kassa	kvitter	kvinna
känguru	känga	lager	lada
koffert	kofta	läger	läge
kolli	kolja	muskel	mussla
konjak	konto	mygel	myra
lego	lera	näver	näve
lotus	loge	paddel	padda
salto	salsa	panter	panna
stereo	stege	puder	puka
sushi	sugga	sadel	saga
tacos	tacka	spader	spade
tango	tanke	smicker	smitta
taxi	taxa	tecken	täcke
telex	tele	tegel	tema
tyfus	tygel	timmer	timme
zombie	sommar	tinner	tinning

(Manuscript received July 15, 2008;
revision accepted for publication July 2, 2009.)