

# The Perception of Word Accents in Swedish

A Response Time Study

Pelle Söderström

BA thesis  
General Linguistics  
Autumn semester 2010

Supervisors:  
Merle Horne  
Mikael Roll

## Acknowledgments

I would like to thank my supervisors, Merle Horne and Mikael Roll for extremely rewarding discussions and suggestions, along with a good dollop of unwavering support, patience and technical help. I would also like to thank Joost van de Weijer for help with the statistical analysis, Johan Frid for his suggestions on stimulus production and Nils Holmberg and everyone else in the Humanities Laboratory who made me feel very welcome. Finally, this study would not have been possible without the help of my lovely participants – thank you!

Last, but not least, a giant pat on the shoulder for Rhiân, who put up with listening to the experimental stimuli far too many times during the production stages. Iawn caws!

## Table of Contents

1. Introduction.....	4
2. Background.....	4
2.1. Swedish word accents .....	4
2.2. Phonological status of the Swedish word accent patterns.....	6
2.3. Previous studies .....	7
3. Hypotheses .....	9
4. Method.....	10
4.1. Experimental design.....	10
4.2. Production of stimuli .....	10
4.3. Procedure.....	12
4.4. Predictions .....	13
5. Results .....	13
6. Discussion.....	16
6.1. Possible effects of musical aptitude on rapid-change acoustic processing .....	17
7. Conclusions and directions for further research.....	18
Appendix.....	23

## 1. Introduction

The purpose of this study is to investigate how the prosodic form of (Central) Swedish word accents interacts with syntax and morphology. More specifically, the study will investigate the relation between verb morphology and word accents, i.e. how a particular stem tone associated with a certain accent pattern is used to predict an upcoming suffix in online speech processing. To achieve this, I will first review work on the theoretical modelling of modern Swedish word accents, and how word accents are assumed to be associated with suffixes in verb (and indeed noun) morphology and then move on to describe a response time experiment conducted to test the hypotheses presented below. The hypotheses will then be discussed in light of the results achieved in the experiment.

A secondary aim of the study is to provide a pilot study and backdrop to a future fMRI (functional Magnetic Resonance Imaging) experiment which will be used to explore the processing of Swedish word accents in the brain. It is important to obtain robust results from the present study, since the paradigms used in the two experiments will be fairly similar. The basic idea is to create a prosodic mismatch between stem tone and suffix, so that the morphological parser breaks down and is forced to reanalyse the utterance. This should then manifest itself as longer response times in a response time paradigm.

## 2. Background

In order to be able to review the different models accounting for the patterning of word accents in Swedish, it is necessary to discuss and also to operationalise the differences between these patterns. I will therefore begin by describing the acoustic properties of Swedish word accents, before moving on to the different accounts of their phonological properties.

### 2.1. Swedish word accents

Word accents are used in most varieties of Swedish, Norwegian and Danish alongside stress as a prosodic category differentiating aspects of word meaning. Although Danish *stød* is phonetically different (expiratory rather than tonal) from the Swedish and Norwegian word accents, it is widely believed to share the same origins diachronically (Bruce 1998). In modern Central Swedish (the general Stockholm variety), the dialect used in this study, there are two distinct word accent patterns, often referred to as *Accent 1* and *Accent 2*, or *acute* and *grave* accent respectively. At this point, it is worth noting that the reason for using the Central Swedish dialect is that some Swedish dialects in the north of the country and in Finland *do not* have word accents. There is a

similar situation in Norway and Denmark, where Northern and Western Norwegian dialects do not have word accents and Southern Danish accents do not have *stød* (Bruce 1998).

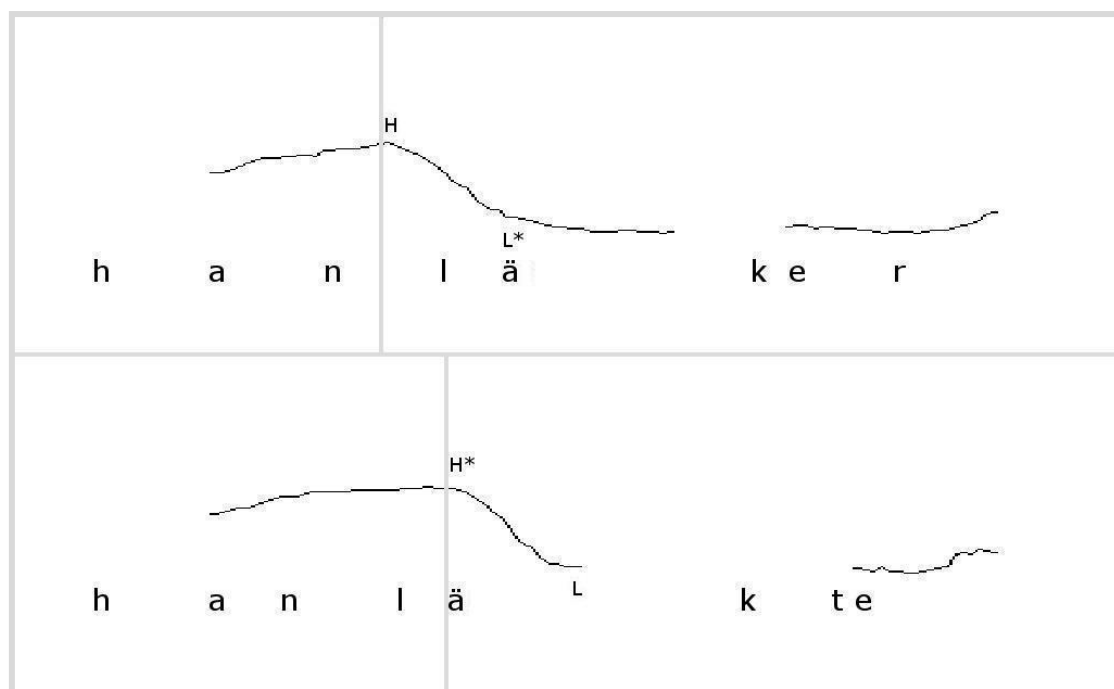


Fig. 1.1. Differences in timing of the word accent  $f_0$  fall between Accent 1 (top, HL\*) and Accent 2 (bottom, H\*L) utterances. The utterances are *HAN läker* (“HE heals”) and *HAN läkte* (“HE healed”) respectively<sup>1</sup>.

Following Bruce’s (1998) analysis, the difference between Accent 1 and Accent 2 can be analysed as one of tonal *timing*, relative to the beginning of the stressed syllable (cf. *lä-* in Fig. 1.1.). Accent 1 is characterised by a tonal gesture which rises to a high tone (H) in the pretonic syllable (the syllable that precedes the stressed syllable) and falls to a low tone at the beginning of the stressed vowel (L\*). Accent 2, on the other hand, is characterised by a high tone (H\*) in the beginning of the stressed vowel which then falls to a low tone through the stressed syllable (L) (ibid. p. 104). Accordingly, the difference can be coded as HL\* (Accent 1) and H\*L (Accent 2). This shows further that the same tonal gesture clearly exists in both accent patterns, but that the difference lies in the timing of the gesture (with respect to the stressed syllable). It is important to note that the sentence accent rise that would indicate a focused verb is absent in the example and in all test material used in this thesis. The only tonal gesture present is thus what Bruce calls the “word accent fall” in his analysis (Bruce 1998:107).

It should be pointed out that some descriptions of the Scandinavian accents include the sentence level focus rise as part of the word accent distinction (e.g. Felder 2009, Felder *et al.*

<sup>1</sup> Tones critically associated with the stressed syllable are indicated by an asterisk (\*).

2009). The difference between the accents then also involves the number of tonal peaks, where Accent 1 words have one peak (focal H) and Accent 2 words have two (word accent H and focal H). This kind of analysis is only relevant, however, when a word occurs in focal position in an utterance. Bruce's analysis removes focus markers from the word level representation so as to isolate the "true" tonal difference between the word accents, namely timing of the tonal gesture (HL\* vs. H\*L).

## 2.2. Phonological status of the Swedish word accent patterns

There are a number of competing accounts of the phonological status of the Scandinavian word accent patterns. A central issue is the question as to whether the accent patterns are associated with certain suffixes in the mental lexicon (cf. Riad 2009) rather than with full word forms. For the purposes of this thesis, this can be expressed as the question as to whether the past tense suffix *-te* is associated with Accent 2 in the lexicon and transfers an Accent 2 pattern to the verb stem to which it is connected when the word is "assembled" for output. Some suffixes, such as present tense *-er* normally co-occur with Accent 1. This has in turn led to suggestions that one of the patterns can be seen as "default", i.e. applied to stems in the absence of a suffix which specifies which accent pattern is to be used. One strong argument for the assumption that Accent 2 is lexically specified and that Accent 1 by contrast is applied "by default", post-lexically, is the fact that speakers of Swedish generalise the Accent 1 pattern onto new unanalysed loan words that enter the language and onto words when speaking foreign languages (Bruce 1998). Lexical marking of Accent 2 has also been explained by diachronic analyses, such as Riad's proposal that the pitch contour of Accent 2 is the result of stress clash resolution in Old Norse, i.e. where secondary stress was removed while the pitch contour remained on the word (Riad 1998). Riad notes that "the crucial novelty of such a system is that it separates pitch information from other stress information, and establishes a phonological tonal tier, which is not a simple mirror of other levels of structure" (ibid. p. 5). According to Riad, this was what led to the lexical accent distinction.

On the other side of the spectrum, Lahiri, Wetterlin and Jönsson-Steiner (2006) have suggested that Accent 1 rather than Accent 2 is lexically specified, and that Accent 2 in that case is "default". The reason for this is apparently that "all thorny problems as well as simple generalisations are dealt with using a simple truism: lexically specified (unpredictable) Accent 1 dominates" (ibid. p. 172). This would apply to "all Scandinavian dialects", in particular Standard East Norwegian and Central Swedish. However, this hypothesis would not be able to explain why unanalysed loan-words receive Accent 1 in Central Swedish (such as *tango*, *bandy* (cf. Riad 2009)). Also, their analysis is based on focal forms of the word accents and thus it is not clear how it

would account for non-focal (basic) forms of the accents. Moreover, the experiment design used by e.g. Felder *et al.* (2009) differs a great deal from the design used in the present thesis. Apart from the fact that only (focal forms of) nouns were used, the experiment was not concerned with suffixes and how they interact with stem tone. Instead, a word fragment (a stressed syllable) with either Accent 1 or Accent 2 was presented auditorily and a word pair was shown on a screen. The task was then to choose which of the word on the screen had been activated by pushing a button on a response box. The participant's response time was measured from the onset of the visually presented word pair on the screen. Response times were found to be the shortest for "specified" Accent 1 words (e.g. words like *hambo*, "hambo" (a traditional Swedish dance)) compared to both "un-specified" Accent 1 words (such as *hummer*, "lobster") and Accent 2 words (such as *hampa*, "hemp" or *humla*, "bumblebee") (ibid.). In this case and context, what is meant by words that are "specified" is that these words "have information on their tonal pattern in their lexical entries" (ibid., p. 1892). Thus, this kind of experiment would not be able to explain how noun or verb morphology interacts with word accent assignment in Swedish, since only full word forms were used. The paradigm used in the present thesis seems more apt for this purpose, since it involves words with clearly analysable stems and suffixes. It is then possible to look closer at the effect suffixes have on word accent assignment.

Recent ERP (event-related potentials<sup>2</sup>) (Roll 2009; Roll, Horne & Lindgren 2010) and other experimental studies (Ambrazaitis 2009) suggest that Accent 2 is lexically specified. These studies will be discussed in more detail below, and their results were used in order to formulate a hypothesis for a response time experiment. If the hypothesis in this experiment is verified, it will be taken as further support for the idea that Accent 2 is lexically specified, as opposed to the hypothesis presented by Lahiri, Wetterlin and Jönsson-Steiner.

### 2.3. Previous studies

Roll (2009) and Roll, Horne and Lindgren (2010) tested Riad's account of Swedish word accents by performing an ERP experiment. They carried out a perception experiment to investigate whether stem tones increased listeners' expectation for certain suffixes.

Roll (2009) and Roll, Horne and Lindgren (2010) aimed to investigate what effects mismatch between stem tone and suffix in nouns has on sentence processing. These effects were compared to those of noun declension mismatches in an ERP study, so that the waveform components resulting from the respective errors could be measured against each other. It is interesting to note that both declension mismatch (e.g. *\*minkor*, with an incorrect plural suffix, cf. *minkar*, "minks")

---

<sup>2</sup> See Appendix for further description of the ERP technique.

and stem tone/suffix mismatch (e.g. *\*minkar*, with an incorrect Accent 1 L on the stem) produced similar positive EEG deflections, namely P600 effects, but in the case of prosodic mismatch, only H-inducing suffixes clashing with a L\* stem tone produced this effect. No such deflection was found for the assumed “lexically unspecified” suffixes (Accent 1) paired with a H\* stem tone (ibid. p. 120). Furthermore, prosodic “mismatch between word accent tone and suffix did not produce any N400 effect” (ibid. p. 119), meaning that the word accent tones most likely are not associated with whole word forms in the mental lexicon, but rather with certain suffixes.

Ambrazaitis (2009) performed a response time experiment aiming to test the effects of canonical prosodic mismatch on the identification of intonation patterns in Swedish. Participants were asked to judge whether the utterances, consisting of an unaccented pronoun and an accented verb in present tense<sup>3</sup>, represented a “confirmation” (with intonation pattern “H-”, a rising focal accent) or a “new information statement” (intonation pattern “H+L-”, a falling sentence accent, “basically the same form as a non-focal Accent 1 pattern”) (ibid. p. 158-161) and response times were logged and analysed. The response time results for Accent 1 words realised with an Accent 2 pattern suggested that no processing difficulty was present in this condition. This could be explained by the fact that “Accent 1 words are not lexically specified to be associated with a specific tonal pattern” (Ambrazaitis 2009:173). The relatively longer response time (and lower identification rates) for the condition where an Accent 2 word was realised with an Accent 1 pattern suggests that this kind of utterance is more difficult to process. This, in turn, can be taken as an indication that Accent 2 is lexically specified and that Accent 1 is “default” since the condition where an Accent 1 word was realised with an Accent 2 pattern did not elicit longer response times (ibid.).

The above results draw a number of very important conclusions: the fact that *only* Accent 2-inducing suffixes preceded by an incorrect Accent 1 L\* tone cause reprocessing of word forms in the brain (i.e. elicit a P600 or longer response times); further, word accent tones do not seem to be associated with whole word forms in the lexicon, but rather with suffixes. These results can then be taken as strong indications that suffixes are indeed specified for Accent 2 in the mental lexicon, and that no such specification exists for Accent 1. It also lends support to the suggestion that a major role of Accent 2 is to aid in morphological processing and to be *connective*, i.e. to signal “that at least one further syllable will follow which belongs to the same word as the stressed one” (Ambrazaitis 2009:44, cf. Bruce 1998). The assumed “default” status of Accent 1 is further strengthened by the fact that H\* tone stems mismatched with suffixes which normally co-

---

<sup>3</sup> e.g. Accent 1: *han lever*, “he is alive” or Accent 2: *jag lovar*, “I promise” (Ambrazaitis 2009:164).



occur with Accent 1 did not elicit any P600 effect indicating restructuring or longer response times compared with “baseline” (Roll 2009, Roll *et al.* 2010:120, Ambrazaitis 2009).

### 3. Hypotheses

The hypothesis that was tested in this study is that **the absence of a high (H\*) stem tone decreases listener expectation that a certain suffix will follow**. Following the discussion in 2.3, this means that we should expect to see slower response times for Accent 2 words being wrongly associated with a low Accent 1 tone (e.g. *\*släppte* with a low tone realised on the stem). If this hypothesis is confirmed, it would lend support to the idea that Accent 2 is lexically specified. It can be expected that the morphological parser would break down and trigger reprocessing in the brain if a lexically specified suffix is preceded by a word stem which does not indicate which type of suffix is to be expected at all. It would also suggest that Accent 1 is indeed the “default” pattern, as is assumed elsewhere (e.g. Bruce 1998, Riad 1998), as compared to the opposite view that Accent 2 is default (e.g. Lahiri *et al.* 2006).

I predict that the response time results obtained in the experiment described below should be in line with the ERP results in Roll (2009) and Roll *et al.* (2010) and the response time results obtained by Ambrazaitis (2009), **with Accent 2 suffixes preceded by mismatched Accent 1 stems** (e.g. *\*läkte* with a L\* Accent 1 stem) **leading to the most reprocessing** (i.e. expected larger hypothetical P600 amplitude **and thus longer response times**). Since the other mismatch condition (H\* Accent 2 stem followed by a lexically unspecified suffix) did not yield *any* P600 effect in Roll (2009) and Roll *et al.*'s (2010) study, the response times for this condition could possibly be the same as baseline (that is, same as non-mismatch “correct” conditions). This would be in line with the results in Ambrazaitis (2009). The hypothesis, then, is that a L\* followed by an Accent 2-inducing suffix will be the most difficult to process since there is no H\* on the stressed vowel to facilitate processing.

The experimental method used in this thesis is different from that of Roll (2009) and Roll *et al.* (2010) in a number of ways. In the present study, participants were asked to judge whether the utterance was in the present or the past tense. This increases the ecological validity of the experiment since there is a better chance of tapping into online linguistic processing using this kind of paradigm compared with asking whether the utterance sounded “wrong” or “OK” for example. The participant is simply forced to make a choice and will probably be more encouraged to “get it right” using a more ecological paradigm. Semantic judgement tasks have the advantage of reflecting more efficiently what word accents are actually used for in online processing. Furthermore, rather than using nouns as in Roll (2009) and Roll *et al.* (2010), verbs

were used as experimental stimuli in the present study, mostly because the conjugation of verbs in Swedish is normally more regular than the declension of nouns.

## 4. Method

### 4.1. Experimental design

The experiment consisted of presenting 80 utterances (subject+verb) to participants via headphones (4 blocks with 20 utterances each). The participants (n=21) were asked to judge whether the utterance represented an action in the present tense or the past tense and the response times were measured and statistically analysed. Both the correct and incorrect conditions were created by digitally splicing stems and suffixes. This process is described in further detail below. There are four kinds of utterances in the experiment:

1. Accent 1, present tense, e.g. *HAN röker*, “HE smokes”, **correct** (n=20).
2. Accent 2, present tense, e.g. *\*HAN röker*, **incorrect** (n=20), mismatched stem+suffix.
3. Accent 2, past tense, e.g. *HAN rökte*, “HE smoked”, **correct** (n=20).
4. Accent 1, past tense, e.g. *\*HAN rökte*, **incorrect** (n=20), mismatched stem+suffix.

The verb was unfocused so that only the word accent pattern is realised on the word (cf. the discussion above, section 2.1). This was made possible by placing narrow focus on the subject. This means that the only tonal difference between the Accent 1 and Accent 2 words is one of accent timing, as is demonstrated in Fig. 1.1. above. The design of the experiment was thus a 2x2 factor design, Accent 1/Accent 2 x past/present, or ACCENT x TENSE.

### 4.2. Production of stimuli

In order to create the stimuli for the response time experiment, 40 phrases were recorded by a 25 year old male speaker of Central Swedish in an anechoic chamber (Humanities Laboratory, Lund University). 20 phrases were of the type *HAN röker* (“HE smokes”, present tense) and 20 were of the type *HAN rökte* (“HE smoked”, past tense). Only verbs with voiceless plosives were chosen for ease of cutting and splicing the stimuli. It is also important to note that in the verbs chosen for use in this study, vowel quality and quantity on the stem did not change with tense (as it does in e.g. *köper/köpte* (“buys”/”bought”). It is therefore only the stem tone which sets them apart. The recorded sound files were then divided into a “stem” part and a “suffix” part by cutting them 10 ms before the explosion phase of the plosive (i.e. *HAN rö-* and *-ker* and *HAN rö-* and *-kte* respectively). See Figure 4.1. for an illustration of the respective splice points.

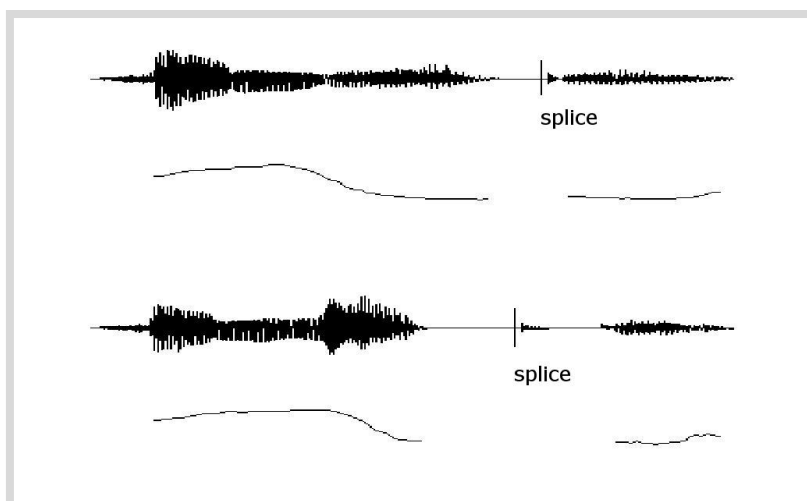


Fig. 4.1. Splice points for the utterances *HAN läker* (top) and *HAN läkte* (bottom). They correspond to, respectively, Conditions PresAcc1 and PastAcc2. Note that the closure phases have already respectively been lengthened and shortened in this figure.

In order to make disambiguation point latency similar across conditions involving different word accents, the silent closure phase before the plosive’s explosion was either lengthened (in Accent 2 stems) or shortened (in Accent 1 stems). With these changes, the stimuli were judged as sounding the most natural as compared to stems with unchanged plosive explosion lengths. All Accent 1 stems were shortened by 60 ms except for the verb stems in *tänka* (“to think”), *bräka* (“to bleat”), *sänka* (“to sink”), *åka* (“to go”), *röka* (“to smoke”) and *steka* (“to fry”) which were reduced by 20 ms. In contrast, the plosive’s closure phase was lengthened by 90 ms in all past tense stems except for *bräka* (“to bleat”) where the closure phase was lengthened by 60 ms. When this was completed, the stems were concatenated with the suffixes using the Praat sound analysis program (Boersma & Weenink 2010).

Condition	Tense	Word accent pattern	Correct
PresAcc1	Present	1 (HL*)	yes
PresAcc2	Present	2 (H*L)	no
PastAcc2	Past	2 (H*L)	yes
PastAcc1	Past	1 (HL*)	no

Table 4.2. The different conditions used in the experiment.

Both the “correct” and “mismatch” conditions (see overview in Table 4.2.) were created in this manner, so that the disambiguation points are the same in conditions involving the same stem, i.e. PresAcc1 and PastAcc1 (i.e. *HAN rö-ker* (Accent 1) and *\*HAN rö-kte* (Accent 1)) and PastAcc2 and PresAcc2 respectively (i.e. *HAN rö-kte* (Accent 2) and *\*HAN rö-ker* (Accent 2)).

Controlling for stem duration across conditions in this way allowed for a more exact interpretation of response time data (measured from the onset of the plosive's explosion phase) and it also helped create stimuli that sounded natural, as judged auditorily by two speakers of Central Swedish.

In the material used for this study, it is interesting to note that the mean  $f_0$  fall from H to L in the two accent patterns seems to resemble an *octave* (12 semitones, re 100 Hz). Mean maximum pitch measured in semitones at H for Accent 1 and Accent 2 words was 12.6, SD=0.4 and 12.9, SD=0.5 respectively. The mean minimum pitch at L for Accent 1 and Accent 2 words was 0.08, SD=0.7 and 1.9, SD=2.1 semitones respectively. The mean differences between H and L for Accent 1 and Accent 2, that is the mean  $f_0$  fall in the two accent patterns, was thus 12.5 and 11.0 respectively. The apparent octave-like distinction between H and L was used in a "pilot" tone decision task described in further detail in section 7. However, this is not meant to be taken as a suggestion that the octave difference is a "universal" feature in speakers of Central Swedish.

### 4.3. Procedure

21 participants participated in the experiment, 11 were male and 10 were female. All participants were right-handed. Mean age was 36.7 years, SD=13.9 (female: 36.3 years, SD=13.7 and male: 37.1 years, SD=14.1). After signing an informed consent form and being told of the details of the experimental procedure, participants were asked if they had any experience of musical training at an advanced level, i.e. if they had been taught to play an instrument in music school and if they had subsequently studied music at university level. Sex, handedness and age of the participant were also noted. All auditory stimuli were presented through headphones. The experiment was performed using E-Prime. A training block was used to acquaint the participants with the paradigm. The task was to press either the 1 or 2 key on the computer keyboard, depending on the tense of the verb in the utterance. Participants were asked to answer as quickly as possible. In 50% of the experiments, pressing the 1 key meant "present tense" and in 50% it meant "past tense", in order to avoid effects of pressing a particular key. 80 utterances were presented in 4 blocks with breaks in between, and response time was logged in E-Prime from the onset of the explosion phase of the plosive marking suffix onset. Between every utterance, a fixation cross appeared on the screen. The disappearance of this fixation cross signaled to the participant that a new utterance was coming up.

A typical test session lasted about 10 minutes. After each experiment session was finished, the participant received a cinema ticket and, if interested, was told about the general details and hypothesis of the experiment.

#### 4.4. Predictions

If it is true that Accent 2 is lexically specified, the mean response times should be significantly longer for condition PastAcc1 (mismatched Accent 1 L\* stem tone followed by an Accent 2-inducing suffix) than for all other conditions, including the other mismatched condition PresAcc2 (see 4.1 for definitions). PresAcc1 and PastAcc2 can be regarded as baseline conditions. Because of the fact that condition PresAcc1 shares disambiguation points with condition PastAcc1 and that condition PastAcc2 shares disambiguation points with condition PresAcc2, it is the mean response times to these pairs of conditions that should be statistically compared with each other. To put it differently: if the hypothesis presented above is true, we should see statistically significant differences in response times to condition PresAcc1 compared with condition PastAcc1, but no significant differences should be expected when condition PastAcc2 is compared with condition PresAcc2. Condition PastAcc1 response times should also be the longest overall, compared to all other conditions.

#### 5. Results

An F1/F2 analysis (mean response time per subject/mean response time per item) was carried out in order to analyse the data collected from the main experiment and this was followed by a *post hoc* Tukey HSD analysis. The F1/F2 analysis showed that there was a significant ACCENT x TENSE interaction ( $F(1,20)=36.7$ ,  $p<0.01$ ;  $F(1,76)=21.8$ ,  $p<0.01$ ). Figure 5.1 shows the interaction plot for the data.

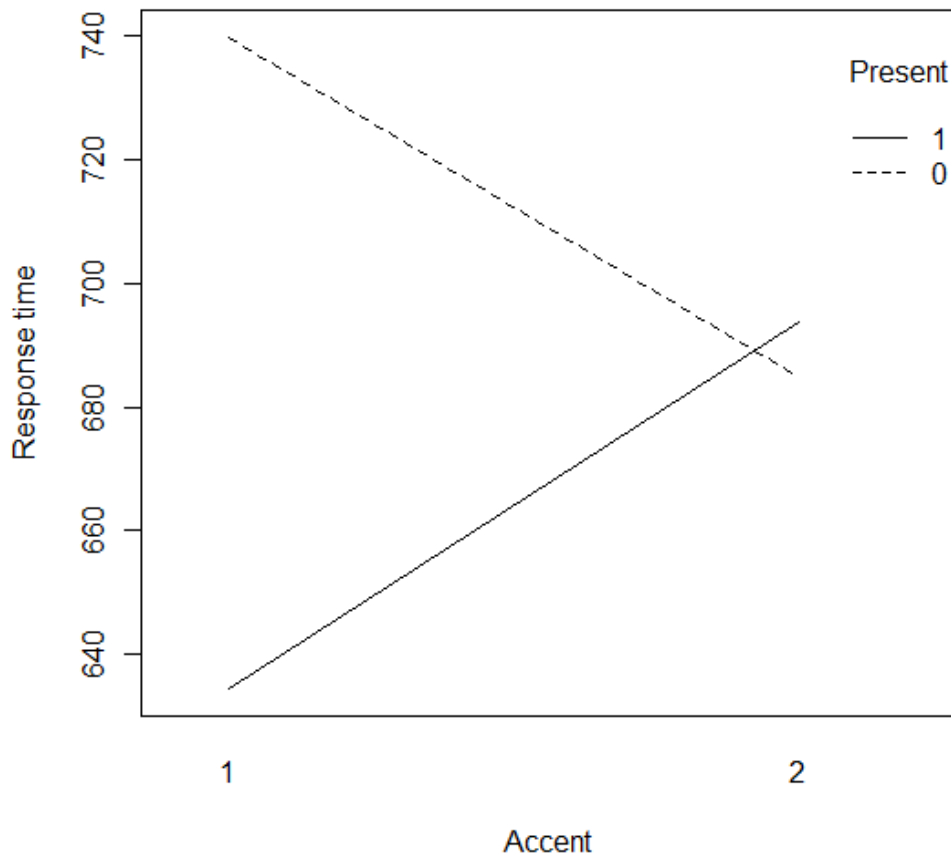


Fig. 5.1. This plot shows the interaction between TENSE and ACCENT. Mean response times are plotted on the y-axis. The black line indicates the response times for verbs with present tense suffixes (Present=1 in the plot) and the dotted line indicates verbs with past tense suffixes (Present=0 in the plot).

The plot in Figure 5.1 shows that response times were the longest when a past tense suffix was preceded by a L\* tone on the stem (condition PastAcc1). Response times were the shortest when a present tense suffixes was preceded by a L\* Accent 1 tone on the stem (condition PresAcc1). A *post hoc* Tukey HSD test showed that there was a significant difference in response time between conditions PresAcc1 and PastAcc1 ( $p < 0.0001$ ) and that there was no significant difference between the conditions PastAcc2 and PresAcc2 ( $p = 0.97$ ). This is in line with the predictions presented in 4.4. Also, there was a significant difference between conditions PresAcc1 and PresAcc2 ( $p = 0.01$ ) and between conditions PastAcc1 and PastAcc2 ( $p < 0.03$ ) The significant mean response time difference between the two conditions PresAcc1 and PastAcc1, measured from suffix onset, was ~105 ms. Figure 5.2 shows the mean response times to the respective conditions, Figure 5.3. plots these means with 95% confidence intervals and Figure 5.4. lists the results of the Tukey HSD test.

Condition	Mean RT	Standard Deviation
PresAcc1	634 ms	278 ms
PresAcc2	694 ms	267 ms
PastAcc2	685 ms	275 ms
PastAcc1	740 ms	329 ms

Fig. 5.2. Mean response times measured from suffix onset along with standard deviation for the four conditions.

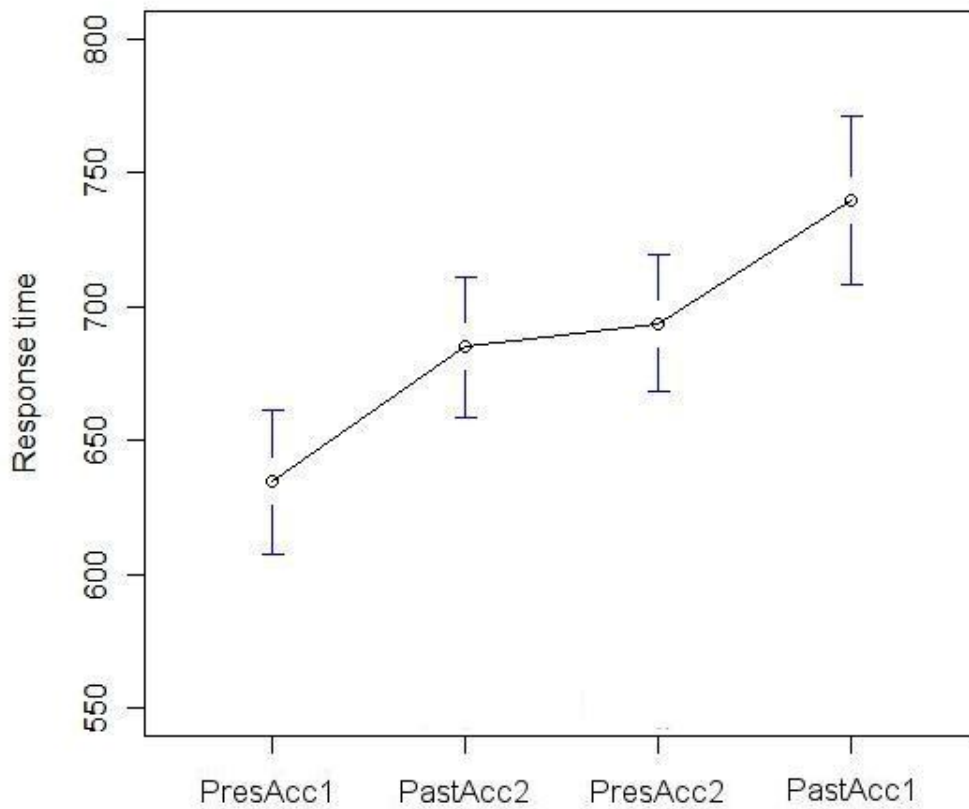


Fig. 5.3. Mean response times for the four conditions plotted with 95% confidence intervals.

Comparison	p-value
PresAcc1 – PastAcc2	<0.05 (*)
PresAcc1 – PresAcc2	0.01 (**)
PresAcc1 – PastAcc1	<0.0001 (***)
PastAcc2 – PresAcc2	0.97
PastAcc2 – PastAcc1	<0.03 (*)
PresAcc2 – PastAcc1	<0.09

Fig. 5.4. p-values from the Tukey HSD.

Another interesting observation is that a H\* Accent 2 tone on the stem seems to both facilitate morphological processing of the past tense suffix *-te* and also to impede processing of the present tense suffix *-er*. However, even when an assumed H-inducing suffix is preceded by a H\* tone on the stem, response times are still longer than when a present tense suffix is preceded by a L\* tone. In other words, the “correct” Accent 1 words seem to be processed quicker than “correct” Accent 2 words (this difference, that between correct conditions PresAcc1 and PastAcc2, was also found to be statistically significant ( $p < 0.05$ )). This significant difference in response times between Accent 1 and Accent 2 was also found by Felder *et al.* (2009). The implications of these findings will be discussed in more detail below.

## 6. Discussion

Following the results of the statistical analysis presented above, the hypothesis that **a past tense suffix preceded by a L\* tone on the stem** (condition PastAcc1) **is more difficult to process** seems to be correct. This is in line with the results obtained by Roll (2009), Roll *et al.* (2010) and Ambrazaitis (2009). This supports the idea that Accent 2 is specified in the mental lexicon on suffixes like *-te* and that the tonal pattern is associated with the stem. When this tonal pattern (H\* on the stem) is absent (or rather replaced), problems in morphological processing occur and when the tonal pattern *is* realised on the stem, processing seems to be facilitated.

The analysis also indicates that the H\* Accent 2 stem tone seems to impede processing of the present tense suffix *-er*. This – along with the fact that even past tense verbs with Accent 2 on the stem take longer to process – means that this tone cannot be said to aid processing in general, but it is possible that it *activates* more word forms compared to the L\* Accent 1 tone. For example, compounds in Swedish take Accent 2 – recall that Accent 2 normally plays a *connective* role – and it is possible that, upon hearing a H\* on a stem, many different things are activated simultaneously<sup>4</sup>. This, in turn, could lead to a general difficulty in processing and longer response times as those shown in the present study. On the other hand, the L\* Accent 1 tone does not seem to activate as many word forms. This is one way of explaining why the present tense Accent 1 words were the easiest to process in the present study, and it could also mean that Accent 1 is applied post-lexically.

However, due to the fact that the H\* Accent 2 stem tone impeded processing of the present tense suffix *-er*, the possibility that both accent patterns are specified in the lexicon cannot be excluded. Furthermore, the results of the present study show that Accent 2 cannot be assumed to

---

<sup>4</sup> For example, it could be hypothesised that a H\* tone on the stem *rök-* facilitates activation of both the past tense suffix *-te* and a compound word like the noun *rökdykare* (a kind of fireman) (or the verb in the present tense *rökdyker*).



be the “default” accent pattern (contra the alternative hypothesis presented by Lahiri *et al.* (2006)) because of the fact that a verb with an assumed Accent 2 inducing suffix preceded by a L\* tone on the stem elicited the longest response times, indicating difficulty in processing caused by the unexpected stem tone.

A possible explanation for the fact that these results deviated in part from the previous studies by Roll (2009) and Roll *et al.* (2010) could be that the experimental method differed in that participants were not asked whether the utterances “sounded” correct or incorrect, but rather whether the utterances were in the present or the past tense. In order to completely compare the results obtained with the ERP results, it would be interesting to perform an ERP experiment using the same paradigm as in the present study.

### 6.1. Possible effects of musical aptitude on rapid-change acoustic processing

In order to be able to effectively process the Swedish word accent distinction and other prosodic structures, speakers are dependent on effective pitch and periodicity processing circuits in the brain. A brain structure which is especially interesting for these purposes is the transverse temporal gyrus, or as it is more commonly called: *Heschl's gyrus* (henceforth HG) after the scientist credited with its earliest description. There have been suggestions that the lateral part of HG in particular acts as a “pitch centre”. In particular, there have been findings that “support the idea that the musical concepts of height and chroma have distinct representations in the human brain” (Griffiths 2003:47), where the pitch chroma corresponds to the “note letter” (e.g. “C”) and the height corresponds to the “distance towards the right on the keyboard” (*ibid.*). Some particular areas surrounding HG were then found to respond to changes in pitch chroma (and not pitch height) and some were found to respond to changes in pitch height (and not pitch chroma) in a PET study. Previous studies have suggested that fast learners of non-native speech sounds had a greater white matter volume in left HG (e.g. Golestani *et al.* 2007). It has further been suggested that white matter anatomy is important when efficient temporal processing is needed, and that grey matter anatomy is more important on larger timescales (i.e. musical processing etc. (*ibid.* p. 580)). However, (linguistic) pitch distinctions were not emphasised in this experiment, and only segmental differences were practiced and analysed (such as the difference between dental and retroflex consonants). It would be interesting to perform a similar experiment with learners of Swedish, and see if the word accent distinction is more easily perceived in learners with greater white matter volume in left HG. Taken together, the indications that HG is important in many aspects of speech sound processing and musical pitch and that white matter volume in HG seems to predict faster phonetic learning, it is possible that relative volume of HG could predict faster processing of pitch differences in Central Swedish for example. However, structures other than

HG have also been implied when acoustic processing in the brain has been studied: in a combined fMRI/EEG experiment, using an MMN (mismatch negativity) paradigm, Schönwiesner *et al.* (2007) further concluded that, along with HG, the posterior superior temporal gyrus, planum temporale and the mid-ventrolateral prefrontal cortex seem to be involved in the automatic processing of acoustic stimuli (Schönwiesner *et al.* 2007:2081). It is clearly not enough, then, to narrow down the area of research to only one structure in the brain, and many other areas will have to be studied further in order to explore the processing of Swedish word accents in the brain.

Musicians (especially professionals) are heavily dependent on efficient pitch processing of different kinds for their work, and this has been shown to be reflected in the relative size of HG components in musicians (professionals and amateurs) as compared to non-musicians (Schneider *et al.* 2005). For this reason, participants in the present experiment were asked about their musical training to see if this had any effect on response time to online pitch processing, both linguistic and musical. It would have been ideal to include a subset of professional musicians in the participant group so that the differences between groups could have been quantified more easily. This is definitely a direction that should be followed in further research into this subject. It is important then to use balanced groups with respect to age and reaction time speed so that participants who process tonal information very efficiently do not lose out because of slow motor response to stimuli related to differences in age. It was for this reason that the mean age of participants was kept fairly low (36.7 years, SD=13.9) and balanced over sexes (female: 36.3 years, SD=13.7 and male: 37.1 years, SD=14.1) in the present experiment.

## 7. Conclusions and directions for further research

The results obtained in the present study suggest that Accent 2 is specified in the mental lexicon and associated with certain suffixes. The H\* tone in Accent 2 did not, however, seem to facilitate general morphological processing as it was found to impede the processing of the present tense suffix *-er*. This could be explained by the fact that the Accent 2 H\* on the stem activates more words forms than the L\* of Accent 1.

A challenge for future research is to look more closely at how the brain handles the Central Swedish word accent distinction. As was discussed in 6.1, given that some brain areas dealing with language and music seem to overlap a great deal, it would definitely be interesting to add musical ability as a factor in future studies. This was partly done in the present study, but the analysis would benefit from a factor which is more easily quantified. Using “professional musicians” and “non-musicians” as two subsets of the participant group would be a more

efficient way of achieving this. If the relative size of an individual's Heschl's gyrus depends on musical training, it could possibly be assumed that we would see differences between the groups when it comes to efficient temporal pitch processing for example. Using functional magnetic resonance imaging, with a similar mismatch paradigm as the one used in the present study, it would perhaps be possible to locate the areas involved in the processing of Swedish word accents. It would be particularly interesting to see if the two accent patterns Accent 1 and Accent 2 are processed differently in the brain. This has been suggested by Roll (2009) where an ERP component, the "P200", was elicited by Accent 2 H\* stems but not by Accent 1 L\* stems. It seems to be a "pre-attentive effect of the perception of pitch movement, processed in the right anterior and temporal cortex" (ibid. p. 157). It would be interesting to try to localise this effect further using fMRI.

In order to look at response times using tonal stimuli, the participants in the main experiment also participated in a different "tonal" pilot experiment. This task consisted of answering whether (the pitch of) a tone was "high" or "low". This test was included in order to look at another kind of response time that does not include any linguistic processing. A short training block presented two different kinds of stimuli (both auditorily and visually: while the "low" tone was playing, text appeared at the lower left edge of the screen and while the "high" tone was playing, text appeared at the upper right edge of the screen). The tones were sine tones (sinusoidal wave-forms), each was 432 ms long. The "low" tone was middle C ( $C_4 \sim 262$  Hz) and the "high" tone was a C one octave higher ( $C_5 \sim 523$  Hz). The reason for using octaves, apart from the fact that they are possibly more accessible to participants with no musical training, was that the H-L fall in the two accent patterns seems to resemble a difference in octave, such as the octave difference of 12 semitones between  $C_4$  and  $C_5$ . Again, the 1 and 2 keys on the computer keyboard were used to perform the task and again in 50% of the experiments, 1 meant "low" and in 50% it meant "high". As in the first experiment, 80 tones were presented in 4 blocks.

No correlation was found between response time and musical training, suggesting that participants who had more experience with musical training did not benefit from this directly during the tone decision task. The mean response times for "low" ( $C_4$ ) and "high" ( $C_5$ ) tones were 534.7 ms and 514.1 ms respectively. However, this difference was not significant ( $p=0.16$ ). The design of this tone decision task could have been made better and more intuitive. Instead of single tones it would be interesting to use a sliding sine tone, or two discrete sine tones, one immediately following the other. The participant would then have to choose whether this tonal gesture represented a "rise" or a "fall". This would place more demands on pitch processing and would be more comparable to what is actually found in linguistic pitch patterns. Naturally, other

things come into play during the perception of pitch patterns in linguistic contexts, and things other than fundamental frequency are also important to the analysis, such as spectral differences etc. However, non-linguistic tests such as this could possibly lead to interesting discoveries if the experimental design is further tested and refined.

## References

- Ambrazaitis, G. (2009): *Nuclear Intonation in Swedish*. Travaux de l'Institut de linguistique de Lund, no 49. Lunds universitet.
- Boersma, P. & Weenink, D. (2010): *Praat (version 5.1.34)* [Software]. [www.praat.org](http://www.praat.org)
- Bruce, G. (1998): *Allmän och svensk prosodi*. Praktisk lingvistik, no 16. Inst. för lingvistik. Lunds universitet.
- Felder, V. (2009): *Neurolinguistic Evidence for the Representation and Processing of Tonal and Segmental Information in the Mental Lexicon*. Ph.D. dissertation, University of Konstanz.
- Felder, V., Jönsson-Steiner, E., Eulitz, C. & Lahiri, A. (2009): Asymmetric processing of lexical tone contrast in Swedish. In: *Attention, Perception & Psychophysics*. N° 71(8), pp. 1890-1899.
- Golestani, N., et al. (2007): Brain Structure Predicts the Learning of Foreign Speech Sounds. In: *Cerebral Cortex*. N° 17, pp. 575-582.
- Griffiths, T. (2003): Functional Imaging of Pitch Analysis. In: *Ann. N.Y. Acad. Sci.* N° 999, pp. 40-49.
- Lahiri, A., Wetterlin, A. & Jönsson-Steiner, E. (2006): Scandinavian lexical tone. In: *Nordic Prosody*. Eds: Bruce, G., Horne, M. Peter Lang. Frankfurt. pp. 167-175.
- Riad, T. (1998): The Origin of Scandinavian Tone Accents. *Diachronica* XV:1, pp.63-98.
- Riad, T. (2009): The Morphological Status of Accent 2 in North Germanic Simplex Forms. In: *Proceedings of Nordic Prosody X*, Helsingfors.
- Roll, M. (2009): *The Neurophysiology of Grammatical Constraints*. Ph.D. dissertation, Lund University.
- Roll, M., Horne, M. & Lindgren, M. (2010): Word accents and morphology – ERPs of Swedish word processing. In: *Brain Research*. Vol. 1330, pp. 114-123.

Schneider, P., Sluming, V., Roberts, N., Scherg, M., Goebel, R., Specht, H., Dosch, H., Bleeck, S., Stippich, C. & Rupp, A. (2005): Structural and Functional Assymetry of Lateral Heschl's Gyrus Reflects Pitch Perception Preference. In: *Nature Neuroscience*. Vol. 8, n° 9, pp. 1241-1247

Schönwiesner, M., Novitski, N., Pakarinen, S., Carlson, S., Tervaniemi, M. & Näätänen, R. (2007): Heschl's Gyrus, Posterior Superior Temporal Gyrus, and Mid-Ventrolateral Prefrontal Cortex Have Different Roles in the Detection of Acoustic Changes. In: *Journal of Neurophysiology*. Vol. 97, pp. 2075-2082.

## Appendix

### An overview of the ERP technique

The ERP technique is based on EEG, electroencephalography, which measures the electrical activity in the brain with the help of electrodes placed on the scalp. The changes in voltage that can be measured in this way “stem partly from postsynaptic electrochemical processes involved in the transmission of information between groups of neurons” (Roll 2009:20). In the case of ERP research, the EEG waveforms are averaged across several trials and time-locked to certain stimuli, such as a word’s onset or other points in time (making it “event-related”). After the waveforms of the different subjects are averaged, “recurrent responses to stimuli in the form of positive (...) or negative potential peaks, referred to as ‘components’, emerge” (ibid. p. 22). One of two interesting components for the purposes of this thesis is the “P600”, an electrically positive EEG wave peaking at around 600 ms. This component has been found to be related to reprocessing in the brain following broken structural (e.g. syntactic, morphological or prosodic) expectations (Roll *et al.* 2010:116). It then follows that the reprocessing difficulty is associated with the amplitude of the P600, so that features of the stimulus that facilitate general processing should decrease the P600 amplitude following a broken expectation in a word or a sentence.

Another interesting ERP component that merits closer description is the “N400”. This negative deflection “decreases as a function of how expected a word is in a particular semantic context, and thus how easy lexical retrieval is” (ibid. p. 115). This means that “if Swedish word accents are associated with whole word forms in the mental lexicon [rather than with particular suffixes PS], incorrect word accents should produce unfamiliar words, increasing the N400” (ibid.).