Neurophysiology of a left-edge boundary tone using natural and edited F0

Roll, Mikael; Horne, Merle; Lindgren, Magnus

Published in:
Proceedings Speech Prosody 2008

2008

Neurophysiology of a left-edge boundary tone using natural and edited F0

Mikael Roll, Merle Horne, and Magnus Lindgren

Department of Linguistics and Department of Psychology
Lund University, Sweden
mikael.roll@ling.lu.se

Abstract

The ERP effects of left-edge, sentence-medial boundary tones in East Swedish were examined using synthesized and non-synthesized F0. The perception of the boundary tone gave rise to an N100 effect, reflecting automated processing of the acoustic features of the stimuli, a broadly distributed P200 effect, similar to what has earlier been seen for sentence-initial pitch accents, and an expectancy negativity, previously observed in response to sentence-medial, expected pitch accents. The N100 effect was clear only for the synthesized stimuli, possibly due to their relatively greater salience. The P200 was only seen in the combined results. The expectancy negativity was present in the combined results as well as for the non-edited stimuli. The F0 synthesis in the environment of voiceless obstruents produced a difference in the F0 level before the disambiguation point of the H boundary tone, which led to an anterior negativity starting around 50 ms before the H.

1. Introduction

Using Event-Related Potentials (ERPs), neurophysiological correlates to right-edge boundaries of Intonational Phrases (IPs), as well as sentence-initial, and sentence-medial expected and unexpected pitch accents have previously been found. Right-edge IP boundaries produce a large positive shift in the ERPs, termed the 'Closure Positive Shift' (CPS) [7]. The first intonation peak in a sentence yields a right-lateralized effect on the amplitude of the P200 [1], an ERP component stemming from the perception of the physical features of an auditory stimulus. Sentence-medial pitch accents elicit an N400 effect if they are unexpected [1]. The N400 is a negative ERP component peaking at 400 ms after critical stimuli that are integrated in a semantic representation [3]. The more difficult a stimulus is to integrate, the larger the N400 amplitude gets. Expected pitch accents produce a frontally distributed negativity at 400–600 ms after stimulus onset, termed 'expectancy negativity' (EN) [1].

In the present study, we examined the ERP effects of the perception of left-edge boundary tones at the beginning of embedded clauses in sentence-medial positions in East Swedish. Embedded clauses introduced by the subordinate conjunction *att* 'that' can be pronounced with or without a H boundary tone in the last syllable of the first prosodic word [2, 5], as shown in example (1), where the subject of the *att*-clause, *kanalen* 'the channel', has a H in its last syllable. The waveform and F0 contour of the example sentence are presented in Fig. 1. A right-edge boundary L associated with the main clause verb *menar* 'think' creates a phrase-final fall. The H associated with the embedded clause subject *kanalen* 'the.channel' marks the left edge of the embedded clause. Embedded clauses with a left-edge boundary tone are both pragmatically and syntactically more similar to main clauses than to subordinate clauses lacking a boundary tone [5, 6], and are therefore referred to as 'embedded main clauses'. Like main clauses, they express their own illocutionary force, assertion in example (1). In canonical Swedish subordinate clauses, sentence adverbs precede the inflected verb linearly, whereas they follow the verb in main clauses. In the *att*-clause in example (1), the sentence adverb *ju* 'of course' follows the verb *sänder* 'sends', just as it would have done in a main clause.

1. *Försvararna menar alltså att kanalen* [H] *sänder*
   The.defenders think thus that the.channel transmits
   ju det på lördag
   of.course that on Saturday

We have previously explored how the left-edge boundary tone affects syntactic processing of the following embedded clause [6]. In the present study, we will analyze the ERPs for the perception of the H boundary tone. An MEG study on the detection of the change from natural voices to natural, F0-edited, or MBROLA-synthesized voices has shown that effects on preattentive perception increase with the degree of synthesis [4].

In order to control for other prosodic features that could influence the processing of the sentences, the difference between the boundary tone conditions and the no boundary tone conditions was created by editing the F0 contour. In half of the stimuli, the original sentence was recorded with a boundary tone, and the F0 level was lowered to create the no boundary tone condition; in the other half of the stimuli, the original sentence had no boundary tone, and the F0 was accordingly raised to create the corresponding boundary tone condition. We will examine the ERP effects of synthesized and non-synthesized stimuli both in combination and separately.

2. Method

19 right-handed students, native speakers of East Swedish, participated in the present study. 6 subjects were excluded from the analysis due to excessively noisy EEG. Of the remaining 13, 7 were female, the age range was 19–32 years, and the mean age was 26.8 years ($s_x = 3.9$).

The subjects were seated in front of a computer screen. The sentences were presented auditorily in loudspeakers placed in front of them. Before and during each sentence, a fixation cross was presented on the computer screen. The subjects were instructed to judge sentence acceptability by pressing one of two buttons.

The stimuli consisted of sentences such as the one presented in (1), with a main clause (*Försvararna menar alltså...* 'The defenders think thus'), and an embedded clause introduced by *att* 'that' (*... att kanalen sänder ju det på lördag* 'that the channel transmits of course that on Saturday'). The waveform and F0 contour of the example sentence are presented in Fig. 1. The same sentence structure, syllable number, and word accent pattern were used for all the sentences. Assertive and nonassertive verbs, such as *menar* 'think' and *hoppas* 'hope', respectively, were each used in half of the main clauses. Half of the embedded clauses contained the sentence adverb *ju* 'of course', and half did not. There were thus 3 factors, Boundary Tone, Verb, and Sentence Adverb, with 2 levels each. 40 sentences per condition rendered 320 sentences in total. The verb and sentence adverb factors have been treated elsewhere [6], and will not be considered here. In the present study, we will consider only the boundary tone factor, and we will analyze results obtained with synthesized and non-synthesized stimuli separately. For the purpose of the study, there will thus be 80 sentences per condition.

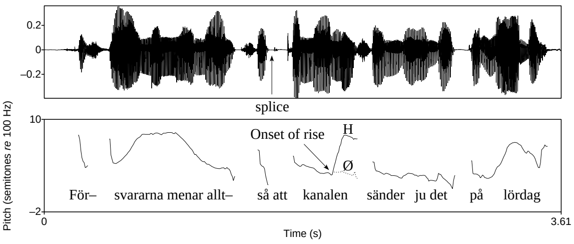

Figure 1: *Waveform and F0 contour of the test sentence* Försvararna menar alltså att kanalen sänder ju det på lördag *'(lit.) The defenders think thus that the channel sends of course that on Saturday'. The embedded clause is shown both with a H boundary tone in the last syllable of its subject* kanalen *'the channel' (solid line), and without one (dotted line, marked with a Ø).*

The sentences were put together from original sentences recorded by a male native speaker of East Swedish in a soundproof, echo-free room at a sampling frequency of 44.1 kHz. Half of the sentences were recorded with a H boundary tone in the last syllable of the embedded clause subject, and half without a boundary tone. The F0 contour was edited in Praat using PSOLA synthesis to create the other half of the embedded clause stimuli. When the original F0 contour contained a H boundary tone, it was lowered to create the corresponding F0 contour without a boundary tone. When there was no H boundary tone, one was created by raising the F0. The F0-editing was based on perceived naturalness of the resulting intonation contour. The F0-peak was timed to align with the onset of the vowel of the last syllable of the embedded clause subject.

The main clauses and embedded clauses were spliced together in the closure phase of the [t] in *att*, just before the occlusion, as illustrated in Fig. 1. The combination of main clauses and embedded clauses was balanced, so that the boundary tone could not be predicted from the main clause intonation pattern.

The test point of the H boundary tone was the part where the F0 contours of the clauses containing a H started differing from those lacking a H. In the case when the last syllable began with a sonorant or voiced obstruent, the point was simply where the rise began in the H sentence. When the last syllable started with a voiceless obstruent, the test point was the voice onset in the syllable.

EEG was recorded using NeuroScan Acquire Software, a 64 channel Quick Cap, and a SynAmps 2 amplifier. Impedance was kept below 5 kΩ. Two horizontal electrodes at the outer canthi of both eyes and two vertical electrodes above and below the left eye were used to record electrooculogram (EOG). The electrodes were referenced to a central reference electrode while recording, and were re-referenced to an average mastoid reference offline. The ground reference was a frontal cap electrode. The sampling rate was 250 Hz, and the cutoff frequencies of the online band-pass filter were 0.05 Hz and 70 Hz. Bad channel signals were replaced offline using spherical spline interpolation with the surrounding electrodes.

Offline EEG data analysis was performed in EEGLAB. The EEG was first filtered with a low-pass filter with a cutoff frequency of 30 Hz, and then divided into epochs of 1100 ms following the disambiguation point of the H boundary tone. A time window of 200 ms preceding sentence presentation was used for baseline correction. The baseline period preceded the epochs with on the average 1958 ms. The distance and the fact that silence in the baseline was compared to audible sentence material were likely to produce large overall amplitudes. The baseline time window was chosen, however, since it could be assumed not to vary systematically with the experimental factors. Significant and marginal results ($p \leq .05$) are reported. All *p*-values are reported with Greenhouse-Geisser correction when applicable. The degrees of freedom are reported with sphericity assumed.

Eye artifacts were compensated for using Independent Component Analysis (ICA). Trials were removed whenever the signal amplitude exceeded ± 100 μV in the baseline or in a time window from 50 ms preceding the epochs to the end of the epochs. Subjects with more than 25% of the trials rejected were discarded from the analysis.

The electrodes were grouped together in 9 Regions of Interest (RoIs), corresponding to 3 anterior/posterior and 3 hemisphere levels: Left Anterior (LA) with electrodes F7, F5, F3, FT7, FC5, FC3, Mid Anterior (MA) with F1, FZ, F2, FC1, FCZ, FC2, Right Anterior (RA) with F4, F6, F8, FC4, FC6, FT8, Left Central (LC) with T7, C5, C3, TP7, CP5, CP3, Mid Central (MC) with C1, CZ, C2, CP1, CPZ, CP2, Right Central (RC) with C4, C6, T8, CP4, CP6, TP8, Left Posterior (LP) with P7, P5, P3, P07, PO5, O1, Mid Posterior (MP) with P1, PZ, P2, PO3, POZ, PO4, OZ, and Right Posterior (RP) with P4, P6, P8, PO6, PO8, O2.

3. Results

3.1. Intonation—synthesized vs. non-synthesized H

The editing of the H boundary tone resulted in an F0 contour that was on the average different from that associated with the naturally produced H. The average F0 contour of edited and non-edited Hs is shown in Fig. 2. For non-synthesized Hs, the lowest point before the rise was 3.24 semitones (st) ($s_x = 0.65$), and the peak was 7.51 st ($s_x = 0.68$), producing a rise of 4.27 st ($s_x = 0.83$). The distance between the measure points was 181 ms ($s_x = 46$), giving a mean slope of 0.0251 st/ms ($s_x = 0.0075$).

For the synthesized Hs, the lowest point before the rise was 3.61 st ($s_x = 0.59$), and the peak was 7.95 st ($s_x = 0.08$), giving a rise of 4.34 st ($s_x = 0.61$). The distance between the measure points was 132 ms ($s_x = 42$), yielding a mean slope of 0.0366 st/ms ($s_x = 0.0136$).

The clauses with a synthesized H thus had an F0 that was 0.37 st higher before the rise ($F_{1,78} = 7.23$, $p < .01$) and had a 0.44 st higher peak ($F_{1,78} = 16.16$, $p < .001$), producing a rise that was non-significantly larger ($F_{1,78} = .15$, $p = .69$) than in the non-edited clauses. The edited clauses had a 48 ms shorter distance between the measuring points ($F_{1,78} = 23.99,\ p < .0001$), yielding a 0.0114 st/ms steeper F0 slope ($F_{1,78} = 21.76,\ p < 0.0001$).

In 32.5% of the embedded clauses, the last syllable of the subject began with a voiceless obstruent. The F0-editing slightly affected the F0 level in the transition to these obstruents. The result was a difference in the F0 contour between the edited H clauses and their corresponding Ø. This difference was located on the average 143 ms ($s_x = 30$) before the point of voice onset after the obstruent, i.e. the disambiguation point of the H. The F0 difference in a 30 ms time window preceding the closure phase of the obstruents was on the average 0.172 st ($F_{1,25} = 96.46,\ p < .001$).

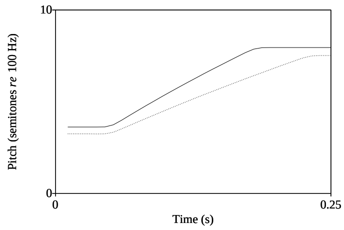

Figure 2: *Averaged F0 from the lowest point in the vowel before the onset of the syllable containing the H left-edge boundary tone, for edited (solid line), and non-edited (dotted line) clauses. The synthesized rise has a higher average F0 and is steeper.*

3.2. ERP data

The combined results for edited and non-edited stimuli after the disambiguation point for the left-edge boundary tone are shown in Fig. 3, and the distribution of the effects, in Fig. 6. Note that the baseline does not show in the figures (see section 2).

There were two effects for H (Tone) as compared to Ø (No tone) in the combined results. At 210–260 ms, there was a positivity that gave rise to a main effect ($F_{1,12} = 4.77,\ p < .05$) (Fig. 6, mid). Between 400 and 600 ms, an increased negativity yielded a Tone×Ant interaction ($F_{2,24} = 11.44,\ p < .01$). Breaking down the interaction revealed an anterior distribution ($F_{1,12} = 7.09,\ p < .05$) (Fig. 6, right).

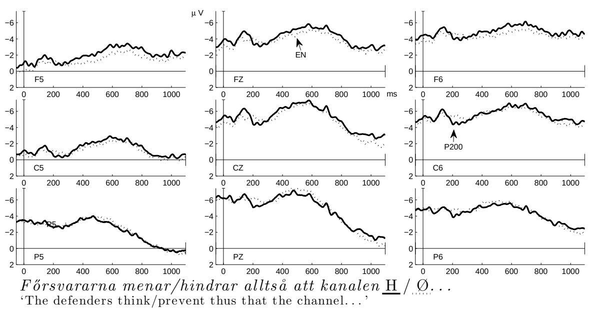

Figure 3: *ERPs for non-edited and edited F0 in embedded clauses together at the disambiguation point of a left-edge boundary tone (H) (solid), or the absence of a tone (Ø) (dotted). There is a broadly distributed positivity between 210–260 ms (P200), as well as an anterior expectancy negativity from 400 to 600 ms (EN).*

The ERPs for the clauses with edited F0 appear in Fig. 4. For the edited stimuli, a negativity around the disambiguation point of the H boundary tone, between −50 and 50 ms, created a Tone×Ant interaction ($F_{1,12} = 11.97,\ p < .001$) showing an anterior distribution ($F_{1,12} = 6.75,\ p < .05$).

Between 130–160 ms after the disambiguation point, there was a broadly distributed negative effect for Tone ($F_{1,12} = 4.97,\ p < .05$) (Topographic distribution in Fig. 6, left).

There were no significant effects between 210–260 ($F = 1.68,\ p = .22$) or 400–600 ms ($F_{1,12} = .51,\ p = .49$ for anterior RoIs).

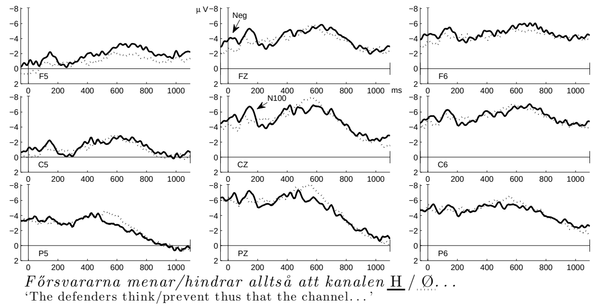

Figure 4: *ERPs for embedded clauses with edited F0 at the disambiguation point of a left-edge boundary tone (H) (solid), or the absence of a boundary tone (Ø) (dotted). An anterior negative effect for H between −50 and 50 ms (Neg), as well as a broadly distributed negativity between 130–160 ms (N100) are visible.*

The ERPs for non-synthesized stimuli are shown in Fig. 5. There were no significant effects at 130–160 ms ($F_{1,12} = .01,\ p = .92$) or 210–260 ms ($F_{1,12} = 1.07,\ p = .32$).

As in the combined results, there was a Tone×Ant interaction ($F_{2,24} = 5.08,\ p < .05$) between 400–600 ms, which was resolved as an anterior ($F_{1,12} = 7.19,\ p < .02$) and central ($F_{1,12} = 4.93,\ p < .05$) effect.

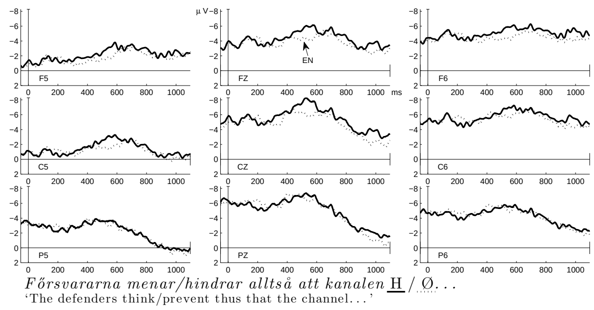

Figure 5: *ERPs for non-edited embedded clauses at the disambiguation point of a left-edge boundary tone (H) (solid), or the absence of a boundary tone (Ø) (dotted). There is a central to anterior expectancy negativity from 400 to 600 ms (EN).*

4. Discussion

We examined the ERP effects of the perception of East Swedish left-edge H boundary tones using synthesized and non-synthesized F0. The rise to the boundary tone created by editing F0 was shorter in duration and had a higher starting point and

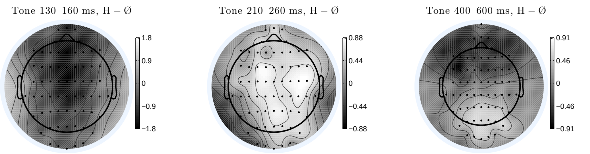

Figure 6: *Topographic difference maps. Left: Negativity at 130–160 ms for the edited stimuli (N100). Mid: Positivity at 210–260 ms in the combined results (P200). Right: Negativity at 400–600 ms in the combined results (EN).*

peak, and a steeper slope than the naturally produced rise. The explanation for this is probably that the duration of the syllable containing the H was naturally lengthened in the non-edited versions due to the production of the rise. No such lengthening was produced in the original recording that served as a base for the synthesized H. The height of the rise did not differ significantly. The editing also created a slight increase in the F0 level in the transition to voiceless obstruents at the onset of the syllable containing the H.

Between -50 and 50 ms around the disambiguation point of the H, there was an increased anteriorly distributed negativity for the synthesized stimuli. We believe that this negativity corresponds to the perception of the difference in the F0 in syllables beginning with voiceless obstruents in the edited clauses.

There were three effects following the disambiguation point of the H boundary tone. The first effect was a broadly distributed negativity between 130–160 ms. This was probably an effect on N100, reflecting perception of the physical features of the stimuli. Similar to the previous negativity, the N100 effect was only significant for the synthesized stimuli. The most obvious explanation for this result is that the synthesized rises were perceptually more salient than the naturally produced rises, since they had a steeper slope, and less variation in the peak frequency.

The second effect, a broad positivity between 210–260 ms, was most likely a P200 effect, similar to what has been found for sentence-initial accents. It was significant in the combined results, but did not reach significance for either the synthesized or the non-synthesized stimuli alone. There may be different reasons why no reliable effect was produced in the separate stimulus sets. In the synthesized stimuli, the difference in the F0 before the disambiguation point might have smeared out the effect. In the non-edited stimuli, on the other hand, the slope of the rise was rather moderate and the peak height varied, which may have yielded a weaker effect. Only the combined effect of both stimulus sets was enough to reach significance.

A later effect, an anteriorly distributed negativity, appeared between 400–600 ms in the combined results, and a similar effect was shown for the non-edited stimuli. We interpret this negativity as an expectancy negativity, found earlier for sentence-medial, expected accents. The boundary tone was expected in the sense that it appeared in a structurally possible location in half of the stimuli. As in the case of the P200 effect, its weakness in the synthesized stimuli might be due to lagging effects of the F0 difference before the disambiguation point. There was no interaction with the edit factor, meaning that the ERPs for the synthesized stimuli also contributed to the combined effect, although the negativity was not strong enough to reach significance separately.

The two early effects, the N100 and the P200, can be attributed to low-level processing of the speech sound. The N100 effect was only present for the edited clauses, probably because the synthesized Hs were perceptually more salient. This result is also in line with earlier findings of stronger preattentive effects in processing of speech with edited F0 as compared to non-edited speech. The later expectancy negativity, on the other hand, was more reliable for the non-synthesized stimuli. The expectancy negativity might reflect higher cognitive processing of the boundary tone.

5. Conclusions

Left-edge H boundary tones in sentence-medial positions display ERP effects that are similar to both sentence-medial and sentence-initial pitch accents. Rises to the boundary tone that had been synthesized were more salient, since they had a steeper slope than the naturally produced rises. The salience was reflected in the ERPs by the fact that the synthesized stimuli showed an N100 effect that was not visible in the natural stimuli. A broadly distributed P200 effect was significant only in the combined results. Its absence in the separate data sets may have been caused by disturbance from the slight F0 difference before the disambiguation point of the H in the synthesized stimuli, and the rather moderate F0 slope and varying peak height in the non-synthesized rises. A later expectancy negativity, interpreted as reflecting higher cognitive processing of the boundary tone, was present in the combined results, and reliable only for non-synthesized clauses separately.

A consequence of the generation of the synthesized Hs was a slight difference in the F0 level in the transition to voiceless obstruents at the onset of the syllable containing the H. This was probably the cause of an anterior negativity in the ERPs from -50 to 50 ms around the disambiguation point of the H.

6. Acknowledgments

The study has been supported by grant 2004-2569 from the Swedish Research Council.

7. References

[1] Heim, S.; Alter, K. 2006. Prosodic pitch accents in language comprehension and production: ERP data and acoustic analyses, *Acta Neurobiologiae Experimentalis*, 66, 55–68.

[2] Horne, M. 1994. Generating prosodic structure for synthesis of Swedish intonation. *Working Papers*, 43, 72–75.

[3] Kutas, M.; Hillyard, S. A. 1980. Reading senseless sentences: Brain potentials reflect semantic anomaly. *Science*, 34, 203–205.

[4] Lattner, S; Maess, B.; Wang, Y.; Schauer, M.; Alter, K; Friederici, A. D. 2003. Dissociation of human and computer voices in the brain: Evidence for a preattentive gestalt-like perception. *Human Brain Mapping*, 20, 13–21.

[5] Roll, M. 2006. Prosodic cues to the syntactic structure of subordinate clauses in Swedish. In Nordic Prosody: Proceedings of the IXth Conference, Lund 2004, G. Bruce and M. Horne (eds.). Frankfurt am Main: Peter Lang, 195–204.

[6] Roll, M.; Horne, M.; Lindgren, M. Submitted. Left-edge boundary tone and main clause verb effects on syntactic processing in Swedish embedded clauses—An ERP study.

[7] Steinhauer, K.; Alter, K.; Friederici, A. D. 1999. Brain potentials indicate immediate use of prosodic cues in natural speech processing. *Nature*, 2(2), 191–196.