The perception of medial stop contrasts in Central Standard Swedish: a pilot study

Pétur Helgason
Dept. of Linguistics & Philology
Uppsala University

Abstract
This paper investigates the perceptual significance of preaspiration in Swedish. Six minimal word pairs were recorded from a male speaker of Central Standard Swedish who had been observed to produce word-medial fortis stops with considerable preaspiration. Each word pair contrasted a sequence of a CV + long fortis stop and a CV + long lenis stop. The CV sequences were then excised from the signal, normalised and presented to 16 listeners, whose task it was to identify the word from which the sequence had been excised. The results show that words with a fortis stop are much easier to identify on the basis of vowel information alone than are words with a lenis stop. This is taken to suggest that preaspiration noise (with a concomitant early voice offset) is highly relevant as a stop cue in Swedish.

Introduction
The word-medial fortis vs. lenis distinction in Swedish has been shown to be more than a simple voiceless vs. voiced contrast (Gobl & Ní Chasaide 1988, Fant et al. 1991, Tronnier 2002, Wretling et al. 2002, Helgason 2002). When producing fortis stops, Swedish speakers tend to initiate glottal abduction before the stop closure is made. As a result, a sequence of a vowel and a fortis stop is often produced with a period of breathy voice and/or aspiration lodged between the vocalic portion and the stop occlusion (cf. Figure 1). This has been referred to as preaspiration. Preaspiration is speaker-dependent and some speakers seldom preaspire, or have very short preaspirations.

A sequence of a vowel and a lenis stop does not contain such preaspiration. Instead, modal voice is typically maintained throughout the production of the vowel, and voicing continues to be maintained during the production of the stop (cf. Figure 1, top right).

Many linguists will agree that preaspiration, as such, is an obscure phonetic detail that is difficult to detect in auditory analysis, and its relevance for perception is thus not immediately clear. Nevertheless, preaspiration has become established as a normative feature of stops in several Scandinavian languages and dialects (Helgason 2002), most notably in Icelandic and Faroese, which suggests that it must have properties that are perceptually salient. The present experiment aims to demonstrate that despite its apparent auditory deficiencies (Bladon 1986), preaspiration does function as a cue for the fortis ~ lenis stop distinction in Swedish.

Method
The experiment involved having Swedish speakers listen to CV stimuli that had been excised from CV + stop sequences, to see whether the vowel in the CV stimulus contained stop type cues. To create the stimuli, recordings were made from a speaker of Central Standard Swedish. The speaker read the following six minimal word pairs in isolation:

- klibba ‘cling to’ ~ klippa ‘cut’
- labba ‘do labs’ ~ lappa ‘mend’
- rädda ‘rescue’ ~ rätta ‘correct (vb.)’
- radda ‘a lot of’ ~ ratta ‘steer’
- snagga ‘cut short’ ~ snäcka ‘talk’

In each of these word pairs, the word on the left contains a lenis stop and the word on the right contains a fortis stop. The speaker who provided the stimulus material had been observed to produce stops with considerable preaspiration. The stop contrast produced by this speaker can be characterised as voiced stop vs. voiceless preaspirated stop.

The stimuli were constructed in the following way. For each word, any information in the signal after the onset of the medial stop closure was deleted. As a simple illustration of this, we can imagine that the stimulus resulting from, for instance, labba is just la-. The vowel durations of the resulting stimuli were then normalised to approximately 130 ms (±3 ms).

The difference between the original recording and the manipulated stimuli is demonstrated in Figure 1. The two upper spectrograms are original productions of two of the test words produced by the stimulus providing speaker.
The two lower spectrograms show the same words with only the initial CV information retained and the vowel duration normalised to 130 ms. These exemplify the stimuli used for the experiment. Stimuli that were obtained from words with a lenis stop will henceforth be referred to as lenis stimuli and those from words with a fortis stop as fortis stimuli.

A total of 16 subjects were recruited for the experiment, 8 male and 8 female, ranging from 27 to 63 years of age. They were from various parts of Sweden, but all lived in Central Sweden (Stockholm or Uppsala) at the time of the experiment, and most spoke a variety of Swedish that would most readily be identified as Central Standard Swedish by native speakers.

For the perception test, a two-alternative forced choice paradigm was used. The subjects listened to the stimuli using headphones and were allowed to repeat the stimuli at will. They indicated their choices in a spreadsheet document. They were discouraged from backtracking and altering previously made choices. However, this was not prevented by the experimental set-up nor by the experimenter, and so some subjects did tend to backtrack and correct their previous choices.

Results

For all subjects pooled, 384 responses were elicited, 192 from fortis stop stimuli and 192 from lenis stop stimuli.

<table>
<thead>
<tr>
<th></th>
<th>Hits</th>
<th>Misses</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fortis stimuli</td>
<td>185</td>
<td>7</td>
<td>192</td>
</tr>
<tr>
<td>Lenis stimuli</td>
<td>144</td>
<td>48</td>
<td>192</td>
</tr>
</tbody>
</table>

Table 1. Response rates for all subjects pooled.

For the 192 fortis stop stimuli, there were 185 hits (i.e. an impending fortis stop was correctly identified) and 7 misses (i.e. an impending fortis was identified as lenis). For the 192 lenis stimuli, 144 were hits and 48 misses. The recognition rate for the fortis stimuli was thus 96.2% and for the lenis stimuli it was 66.6%.

Figure 1. Spectrograms of the original productions of the words radda (top left) and ratta (top right). The stimuli created are shown below: ra- from radda (bottom left); and ra- from ratta (bottom right).
The majority of subjects had a perfect hit rate for the fortis stimuli (see Figure 2). Thus 11 out of 16 subjects always identified the fortis stimuli correctly as coming originally from syllables containing fortis stops. Four subjects had one miss, and thus identified 11 out of 12 fortis stimuli correctly. The remaining subject had three misses for the fortis stimuli.

Generally, the subjects had lower hit rates for lenis stimuli than the fortis ones. Four subjects had a perfect hit rate for the lenes. The remaining subjects had varying hit rates, ranging from 3 to 11 hits for the 12 stimuli.

Discussion

The results show that, on the basis of information from the preceding vowel alone, it is easier for listeners to identify an impending fortis stop than an impending lenis stop. The vowels in the fortis stop stimuli in this experiment contain fairly reliable information that tells the listener that a fortis stop is impending. Such reliable information is not present in the lenis stimuli and therefore listeners find it more difficult to identify an impending lenis stop.

As described in the method section, the stimuli for the experiment were obtained from a speaker who preaspirates his fortis stops quite markedly. It is suggested here that preaspiration, manifested as aspiration noise and/or breathlessness, is largely responsible for prompting the perception of an impending fortis stop, to the exclusion of a lenis stop percept.

The observed difference in hit rates has a direct correspondence in the prevailing norms for stops production in Swedish. Some speakers of Swedish seem to use preaspiration more or less as an integrated feature of their fortis stop production while others tend to use it more sporadically. Speakers who preaspirate sporadically sometimes have preaspiration before fortis stops and sometimes not. As a result, those speakers often have modal voice throughout a vowel preceding a fortis stop. Therefore, having modal voicing throughout the vowel is not a useful stop type indicator—the impending stop may be either fortis or lenis. The presence of preaspiration, however, is unambiguous and can only indicate an impending fortis stop. Thus, when present, preaspiration is a reliable cue for identifying an impending fortis stop.

The fact that some subjects seem to be able to identify an impending (but absent) lenis stop may be a result of the experimental set-up. These subjects may be extrapolating from the fact that preaspirations are so obviously indicative of an impending fortis stop that a lack of preaspiration must be indicative of an impending lenis stop. If this hypothesis is correct, the correct identification of impending lenis stops is the result of a top-down processing rather than the information available in the signal. Thus the subjects have learned to identify an impending lenis stop on the basis of fortis stop identification.

Although it now seems clear that the preaspiration phenomenon triggers a fortis percept for Swedish listeners, it is less clear which acoustic aspect of preaspiration is most salient for perception. For voiceless preaspirations, listeners are likely to make use of the friction noise itself. However, the early cessation of voicing as such makes for a shorter vocalic portion in the utterance, which may provide a durational cue. Thus listeners may also be relying on the shortness of the vocalic portion in the fortis stimuli. A more complex experimental set-up might be devised whereby one could assess the relative contributions of these two factors, aspiration noise and vocalic duration. This might be a fruitful avenue to explore given the apparent tendency in languages for vowels before voiceless consonants to be shorter than vowels before voiced consonants (see e.g. Chen 1970).
For preaspirations that are wholly or partially voiced, the picture becomes even more complex. There is every reason to consider the breathy voiced portion of a vowel-to-stop transition to be part of the preaspiration percept. Wayland & Jongman (2003) list a number of candidates that may be relevant for the perception of breathiness. These include the relative amplitude of the first harmonic ($H_1$), additive noise, spectral tilt, tracheal coupling, fundamental frequency, and acoustic intensity.

![Figure 4. Spectra taken at a point 30 ms prior to stop closure in the words klibba (left) and klippa (right), showing the amplitude of $H_1$ relative to $H_2$.](image)

The first two of these candidates seem to be the most relevant for perception. The relative amplitude of $H_1$ (measured, for example, as the difference in SPL between $H_2$ and $H_1$) has been identified as a consistent acoustic correlate of breathiness. Almost all the fortis stimuli in the present experiment had a period of breathiness during the preaspiration phase. In these breathy portions, $H_1 - H_2$ is consistently higher than at comparable places in the lenis stimuli (see Figure 4). From the point of view of perception, several studies from a number of languages have indicated that the relative amplitude of $H_1$ is a fairly robust cue for breathiness (see overview in Wayland & Jongman 2003).

Experiments assessing the salience of aspiration noise in the perception of breathy voiced vowels have been less conclusive. Some experiments have indicated that there is a fairly strong relationship between additive noise (or lack of periodicity) and perceived breathiness while others have failed to show such a relationship (cf. discussion in Wayland & Jongman 2003). However, there is really no reason to exclude any of these possible cues when discussing the breathiness percept. The amplitude of the first harmonic, noise, tilt, tracheal coupling, etc. are all likely to contribute to the breathiness percept in differing degrees.

This experiment has shown that listeners can identify impending fortis stops quite reliably on the basis of the preceding vowel alone, at least for the type of stimuli used in this experiment. Similar identification of impending lenis vowels is far less successful. The simplest interpretation of this is that there is information in the fortis stimuli that allows those stimuli to be unambiguously identified as fortis. Such disambiguating information seems not to be present in the lenis stimuli, and hence the lenis stimuli are variously interpreted as either fortis or lenis. As it is demonstrably present in all the fortis stimuli, preaspiration is the most obvious candidate for providing this disambiguating information.

References


