PHONOLOGICAL NORMS IN FAROESE SPEECH SYNTHESIS

Pétur Helgason* & Sjúður Gullbein*

*Department of Linguistics, Stockholm University — petur@ling.su.se
*Department of Faroese, University of the Faroe Islands — sg@kallnet.fo

ABSTRACT
In developing Faroese text-to-speech synthesis, the choice of phonological norms has proved problematic. In particular, the literature on Faroese phonology offered contradictory accounts of the realization of word-medial postvocalic stops. Spontaneous speech data from 4 speakers of the Tórshavn dialect were recorded, and their stop production was analyzed. It emerged that there were detailed differences between the 4 speakers in the stop production patterns. Since there seems to be no single “correct” phonological representation of these word-medial stops in the Tórshavn dialect of Faroese, our approach has been eclectic, in that we have chosen those phonological norms for the synthesis which we have judged optimal for the user.

1. INTRODUCTION
Faroese is spoken by approximately 45,000 people on the Faroe Islands in the Atlantic. It belongs to the West-Slavonic branch of the Germanic language family, and has close linguistic affinities with Icelandic in the west and the dialects of Western Norway in the east. There are a number of dialects which differ considerably. The dialect under consideration here is the Tórshavn dialect.

There are two contrasting stop series in Tórshavn Faroese, which will here be referred to as fortis and lenis. The use of these terms provides a way of distinguishing between the two stop series which have varying exponents (i.e. realizations). In traditional accounts, the fortis vs. lenis contrast in word-initial position is described as voiceless postaspirated vs. voiceless aspirated. In word-medial, postvocalic position the contrast is described as voiceless preaspirated vs. voiceless unaspirated. However, the authors have noted that word-initially, the lenis stops are often voiced, and that stops are very often produced as fricatives without any stop closure. Thus, the terms fortis and lenis should not be taken as a statement about the phonetic exponence of the two stop series.

In some postvocalic contexts, specifically after certain long vowels, speakers do not have a fortis vs. lenis distinction. Stops in such context are unaspirated and can be categorized as lenis stops. However, there has not been a consensus as to which contexts induce such a collapse. Earlier accounts (e.g. [1] and [2]), which do not specifically describe the Tórshavn dialect, claim that a fortis vs. lenis distinction is only made when the vowel preceding the stop is a diphthong that ends as a high vowel, [iː] or [uː] (i.e. after [iːi], [uːu], [œi], [eː], [iːr], [æː], [iːi] and [œː]). Thus, the fortis vs. lenis distinction has collapsed in all other contexts (i.e. after [œː], [eː], [œːi], [œː], [œːi], [iː] and [æː]).

More recent accounts (e.g. [3] and [4]) maintain an almost reverse description. A fortis vs. lenis distinction is made only if the preceding vowel ends as a non-high vowel (i.e. after [œː], [œːe], [œːi], [œː], [œː] and [æː]). In other contexts, the distinction has collapsed (i.e. after [œːi], [œːi], [œː], [œːi] and [œː]). This type of distribution of preaspiration in Faroese will henceforth be referred to as the “Modern School Distribution.” These recent accounts, however, are not specific about whether preaspiration in the Tórshavn dialect has this type of distribution.

The development of the Faroese synthesis required a definitive answer as to how these stops were to be produced. The aim of this study is twofold. First, we want to examine how well the presence vs. absence of preaspiration corresponds to the fortis vs. lenis contrast in postvocalic word-medial position as stipulated by the Modern School Distribution. Second, should the Modern School Distribution fail to describe the distribution of preaspiration in Tórshavn Faroese, we wanted to clarify what were the phonological norms for stops in the Tórshavn dialect, especially those following long vowels.

While it would have been advantageous in many ways to use read speech for this investigation, we had reason to think that this might be problematic. Regarding the elicitation of phonological norms, we had noted that it was often difficult for informants to cite a “canonic form” for a given Faroese word. In the case of preaspiration, the contexts where a stop follows a long vowel were particularly problematic. Informants either disagreed on whether a particular stop was preaspirated, or were uncertain as to their intuitions. We had even noted, on a few occasions, that informants altered their production from unaspirated to preaspirated when repeating a word spoken by an interpreter, apparently influenced by the interviewer’s pre-aspiration. For these reasons, we suspected that orthographic representations, which usually “favor” fortis variants, might influence the speakers choice of stop variants. We believed that spontaneous speech data would better capture the speakers’ “intuitive” stop production.

2. METHOD
The Faroese data come from four informants, two male (SG and EI) and two female (MS and BJ), all speakers of the Tórshavn dialect.
dialect. Spontaneous speech interaction between the informants was elicited using map tasks. The informants were recorded in pairs on separate channels. EI and MS formed one pair, SG and BJ another. The data were recorded on a DAT recorder, using directional microphones to minimize overlap between channels. This overlap was further reduced by placing the informants so that they faced away from one another. The microphones were mounted on a headset so that they were placed approximately 2.5 cm out and to the side of the corner of the informant’s mouth. The recordings were made on location in the Faroes. A sound-treated room was not available for the recording sessions, and therefore some echo is present in the recordings.

The total recording time was 20-35 minutes per subject. The amount of material analyzed for each subject varies. For EI, approximately 13 minutes of uninterrupted speech, comprising 2286 words, were analyzed; for MS 14 minutes, 2124 words; for SG 11 minutes, 1887 words; and for BJ 16 minutes, 2845 words. The material was transcribed by one of the authors, a native speaker of Tórshavn Faroese.

In labeling, the same acoustic events were identified for all stops, irrespective of their orthographic representation or historical origin. Thus no assumptions were made as to their phonological status. The relevant events labeled were:

- Vowel onset
- Modal voice offset
- Voice offset
- Onset of stop closure
- Onset of stop release

Sometimes, preaspiration manifests itself only as a breathiness in the vowel preceding a stop [5]. This entails, that a pre-aspiration can be quite audible, even if voice offset does not occur at all before the stop closure is reached. Therefore, we prefer to use modal voice offset time, rather than simply voice offset time, as a measure for the onset of preaspiration.

3. RESULTS

In the following analysis, the stops and the contexts in which they occur are categorized on the basis of four criteria. First, the stops are divided into fortis and lenis stops in accordance with the Modern School Distribution (cf. sect. 1). Second, the stops occur in two main syllable types, VC and VC: / VCC. In simplified terms one can say that in VC syllables the vowel is long and in VC:/VCC syllables the vowel is short, but the ratio between vowel and consonant duration is probably a better determinant of the quantity type than the length of the vowel per se. Third, the second consonant in VCC syllables can be voiced or voiceless. And fourth, the stops can occur in content words or in function words.

To make it easier to refer to the relevant combinations of these variables, a specific word is used to represent and exemplify a given category. Not all possible combinations of these variables are represented, since many do not actually occur in the data, or occur very infrequently. Also, for some categories, no distinction is made between content and function words. The exemplifying words are listed in Table 1.

<table>
<thead>
<tr>
<th>Exemplifying word</th>
<th>Stop category</th>
<th>Syllable type</th>
<th>Word type</th>
</tr>
</thead>
<tbody>
<tr>
<td>átt[a]</td>
<td>fortis</td>
<td>VC:</td>
<td>cont</td>
</tr>
<tr>
<td>oddur [sto:os]</td>
<td>lenis</td>
<td>VC:</td>
<td>cont/func</td>
</tr>
<tr>
<td>hatta [li:htə]</td>
<td>fortis</td>
<td>VC:</td>
<td>func</td>
</tr>
<tr>
<td>vatnið [vətni:θ]</td>
<td>fortis</td>
<td>VCC</td>
<td>cont</td>
</tr>
<tr>
<td>fjallið [fjostla]</td>
<td>lenis</td>
<td>VCC</td>
<td>cont/func</td>
</tr>
<tr>
<td>seks [seks]</td>
<td>fortis</td>
<td>VCC</td>
<td>cont</td>
</tr>
<tr>
<td>fótar [fɔ:tar]</td>
<td>lenis</td>
<td>VcC</td>
<td>cont/func</td>
</tr>
<tr>
<td>matur [matur]</td>
<td>fortis</td>
<td>VcC</td>
<td>cont</td>
</tr>
<tr>
<td>hvat [kvætθ]</td>
<td>'what'</td>
<td>fortis</td>
<td>VcC</td>
</tr>
<tr>
<td>radari [raðar]</td>
<td>lenis</td>
<td>VcC</td>
<td>cont</td>
</tr>
</tbody>
</table>

The syllable type is indicated in the spelling of the exemplifying word. When the stop is geminate or followed by another consonant, the syllable type is always VC: or VCC.

The word types which have fortis stops are exemplified by the words átt[a], hatta, vatnið, seks, matur and hvat. The lenis stop word types are oddur, fjallið, fótar and radari. As for the lenis stop in fjallið, it can be noted that stops in such clusters predominantly come from Old Norse ll and rn, i.e. geminate dental or alveolar laterals and nasals which have undergone “pre-stopping” and become [tl] and [tn] respectively [6]. The Old Norse cluster rn, which usually becomes [tn], is also a source for such clusters.

Syllables with primary and secondary lexical stress are grouped together in this analysis. Thus the analysis does not, for example, make a distinction between the ‘s in báðar and seglúðar. With regard to the r, both are treated as matur type words (i.e. content words where a fortis stop occurs in a VcC syllable). Also, although the stops in the exemplifying words are usually intervocalic, the words occurring in the data are not necessarily followed by a vowel. For example, the word nít ‘middle’ may occur utterance-finally or before a word starting with a consonant, but it still belongs to the átt[a] word type.

Some stops that occur in a word-medial, postvocalic context were excluded from the analysis. This applies to stops that occur after a morpheme boundary (e.g. skattakista ‘treasure chest’) and stops that are in the onset of a stressed syllable (e.g. mobítt ‘mobile’). Both types of context trigger the production of a postaspirated stop, i.e. the type of stop that occurs word-initially rather than medially. Also, the neuter definite article eitt, the preposition and infinitive marker at and the negation ikk[i] were not included in this analysis, since they are rarely produced with a stop closure. Similarly, the word veit in the phrase eit veit ikk[i] ‘I don’t know’ was excluded since it is very rarely produced with a full stop closure.

A total of 1356 word-medial stops were analyzed for the four informants. Of these, 195 cases were discarded for various reasons. In a further 318 cases the informant produced a voiced or voiceless fricative without a full stop closure. No measure-
ments of preaspiration were obtained for these cases. In 843 instances the stop was produced with a complete stop closure. For these cases, preaspiration duration was measured as described in section 2. In 534 of these cases, no preaspiration was registered. The remaining 309 instances were produced with varying degrees of preaspiration, ranging from approximately 5-160 ms.

In the following discussion, stops produced with less than 15 ms of preaspiration will be referred to as unaspirated and stops with more than 15 ms will be referred to as preaspirated. Setting a “preaspiration limit” at 15 ms reflects an unavoidably arbitrary choice on part of the authors. One of the aims of this investigation is to determine how well the predefined linguistic categories of fortis and lenis stops predict the degree of preaspiration in stop production. Unlike the phonological, binary fortis vs. lenis division, the degree of preaspiration is a continuous variable that provides no natural separation line between preaspirated and unaspirated stops. Thus, when preaspiration duration is intermediate, and the stop is neither clearly unaspirated nor clearly preaspirated, we have no way of determining whether the speaker intended to produce a fortis or a lenis stop. Either we omit such cases from our analysis, or we subjectively set a division line between unaspirated and preaspirated. We have chosen the latter approach.

Figure 1 shows the number of occurrences of different types of stop production, where the results for all four informants are pooled. The length of each bar (i.e. the x-axis) indicates the number of observations made for the different word categories. Each bar is divided into three segments. The leftmost segment indicates the number of unaspirated stops, i.e. stops produced with less than 15 ms of preaspiration. The middle segment indicates the number of stops produced as fricatives without a stop closure. The rightmost segment indicates the number of preaspirated stops, i.e. stops produced with more than 15 ms of preaspiration.

Collectively for lenis stop word types, i.e. oddur, fjallid, fótur and radari, only 20 out of 552 instances are preaspirated. When preaspiration does occur in these cases, it is characterized by voiced friction. Usually, these preaspirations can be attributed to the fact that a close vowel preceding the stop creates a narrow escape channel that creates a friction noise and induces voicelessness. For the fortis stop word types áttu and vatnís, the results are reversed. Only 23 out of 280 instances are unaspirated. For all these word types there is obviously a very strong correspondence between unaspirated and lenis on the one hand and between preaspirated and fortis on the other.

For the remaining fortis stop word types, natur, hatta, hvat and seks, the fortis vs. lenis distinction is less clear cut. For the word type natur, informant SG produces a preaspirated stop in 14 cases, an unaspirated stop in 4 cases and a fricative in 13 cases (see Table 2). Despite the numerous fricative realizations, his stops in natur type words can be characterized as preaspirated.

Table 2. Three types of stop production for the word type natur.

<table>
<thead>
<tr>
<th></th>
<th>SG</th>
<th>BJ</th>
<th>EI</th>
<th>MS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preasp. &lt; 15 ms</td>
<td>4</td>
<td>15</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>Fricative</td>
<td>13</td>
<td>15</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Preasp. &gt; 15ms</td>
<td>14</td>
<td>14</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>31</td>
<td>44</td>
<td>18</td>
<td>17</td>
</tr>
</tbody>
</table>

Informant BJ also tends to preaspirate stops in natur type words. However, following [e], stops are almost never preaspirated in her speech. Thus she produces eta ‘to eat,’ tekur ‘takes (vb.),’ and sentimetur ‘centimeter’ without preaspiration before the relevant stop. Table 2 is therefore somewhat misleading with regard to her data. In fact, 12 out of the 15 unaspirated stops occur after an [e]. Of the 14 preaspirated instances, there is only one that occurs after an [e]. It therefore seems that in BJ’s speech, words with a sequence of [e] + stop should be regarded as fótur type words rather than natur type words.

The two remaining informants tend not to preaspirate at all in natur type words. Thus EI produces an unaspirated stop in 15 cases, a fricative in 3 cases, but has no instance of a preaspirated stop. Informant MS produces an unaspirated stop in 15 cases, a fricative in one case, and a preaspirated stop in one case.

The informants also differ in their production of hatta type words. Informant SG tends not to preaspirate stops in such words. Only 3 cases are preaspirated, while 13 cases unaspirated and 35 are produced as a fricative (see Table 3). Thus, fricatives are, in fact, the dominant pattern for SG’s production of stops in this word type.

Table 3. Three types of stop production for the word type hatta.

<table>
<thead>
<tr>
<th></th>
<th>SG</th>
<th>BJ</th>
<th>EI</th>
<th>MS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preasp. &lt; 15 ms</td>
<td>13</td>
<td>12</td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>Fricative</td>
<td>35</td>
<td>19</td>
<td>44</td>
<td>17</td>
</tr>
<tr>
<td>Preasp. &gt; 15ms</td>
<td>3</td>
<td>0</td>
<td>10</td>
<td>8</td>
</tr>
<tr>
<td>Total</td>
<td>51</td>
<td>31</td>
<td>58</td>
<td>33</td>
</tr>
</tbody>
</table>
Stops in *hatta* type words do not seem to be preaspirated in BJ’s speech either. There are 12 cases without preaspiration, and 19 cases a fricative production, but no preaspirated cases. Again, fricatives are the dominant pattern in the production of stops for this word type.

Informant EI produces a preaspirated stop in *hatta* type words in 10 cases, an unaspirated stop in 4 cases and a fricative in 44 cases. Thus he has a definite tendency to preaspirate the stop in such words. Nevertheless, the fricative variant is by far the most frequent.

Finally, informant MS has an equal distribution between preaspirated and unaspirated variants in *hatta* type words, 8 instances of each. There are also 17 instances of fricative production, so again the fricative is the most frequently occurring variant.

As regards *hvat* and seks, there are not enough occurrences to allow any accurate statements as whether they tend to be preaspirated or unaspirated. Still, for the seks word type, there are some indications that informants behave differently. It is possible that BJ lacks preaspiration in ks clusters. There are 6 occurrences of seks type words in BJ’s data, 4 instances with a ks cluster and 2 instances with a kt cluster. All 4 ks clusters lack preaspiration, while both kt clusters are preaspirated. For MS’s seks word type, 4 out of 6 instances are preaspirated and the remaining 2 are unaspirated. The preaspirated instances include 3 ks clusters.

4. DISCUSSION & CONCLUSIONS

The aim of this investigation was to bring some clarity into what are the norms for producing postvocalic, word-medial stops in Tórshavn Faroese, for the purposes of creating Faroese speech synthesis. We found that recent descriptions of the phonological distribution of preaspiration in Faroese were, by and large, consistent with the data. In VC: syllables, and in VCC syllables where the second consonant is voiced, the fortis series is generally realized as a preaspirated stop or a fricative, while lenis series is generally realized as an unaspirated stop.

However, function words with a VC: structure (*hatta* type words) have an unclear status. A fricative is the predominant realization for stops in such words for all informants. However, when a stop closure is produced, the informants differ. One informant (BJ) never preaspirates, one seldom preaspirates (SG), one has one half of the instances preaspirated and the other unaspirated (MS), and one has predominantly preaspirated variants (EI). From the point of view of the synthesis, we find that we are free to choose either variant, preaspirated or unaspirated, for the production of such words. We have opted for the preaspirated variant.

For words with a VcC syllable, the informants also behave differently. Two of the informants (EI and MS) never preaspirate stops in such syllables. One informants (SG) only preaspirates stops in such syllables if the preceding vowel ends as a non high vowel (i.e. after [ɔi], [ɛi], [ɛ], [ə] and [xl]). This pattern is repeated for the last informant (BJ) with the exception, that stops are not preaspirated after [ɛ] in her speech. Again, from the point of view of the synthesis, we can choose which type of distribution preaspiration has in VcC syllables. In this case, we have opted for the type of distribution that the informant SG exhibits.

Fricative realizations are usual for both fortis and lenis stops in many contexts. While this is of interest for the description of phonetic variation in Faroese stops, we find that producing stops as fricatives in the synthesis would be misguided. One of the aims of the synthesis is making the pronunciation “transparent,” i.e. making it possible for the user to deduce the spelling from the pronunciation, so choosing a phonetic variant that is ambiguous, like the fricative in this case, would not be optimal for the user, however “phonetically correct.”

Given that we have only examined 4 informants, it seems highly probable that if we were to examine more informants, further types of stop production patterns would emerge. Phonological variation of this type is difficult to accommodate in speech synthesis, and it is probably a fairly unusual problem in synthesis development. However, it may be expected when dealing with a language that has many dialects and lacks a standardized pronunciation.

5. REFERENCES


