Good timing: Place-dependent voice onset time in ejective stops

Ian Maddieson
Department of Linguistics
University of California, Berkeley
ianm@socrates.berkeley.edu

Abstract
Voice onset time after voiceless unaspirated stops demonstrates a dependence on place of articulation, most reliably being shorter for labial and coronal than for velar stops. Some of the proposed explanations for this pattern suggest that a parallel dependence is not be expected for aspirated or ejective stops. However, similar patterns do occur with both aspirated and unaspirated stops. Cho and Ladefoged [1] have suggested that ejectives do not follow the same trend, but they had little data on bilabial ejectives to compare with more plentiful data on velars. This paper contributes more material to this debate with expanded data on Yapese and the first published material on ejective VOT in Nez Perce. The results suggest that ejectives have a similar pattern to plosives and that therefore a unified explanation for all three types of stops should be sought.

1. Introduction
It is well known that after the release of a prevocalic voiceless unaspirated stop the time that elapses before voicing begins for the vowel shows dependence on the place of articulation of the stop. The voice onset time (VOT) is quite reliably shorter after a bilabial ([p]) than after a velar stop ([k]), with coronal stops often being intermediate and almost always shorter than velars (see, for example, [2, 3]). One proposed explanation for the patterns seen is that the rate of aperture increase in the releasing gesture differs for different articulators and locations. For example, Stevens [4] estimates that at the release of a labial stop the cross-sectional area is increasing at about 100 cm²/s, in an alveolar stop it is increasing at about 50 cm²/s, and in a velar stop at about 25 cm²/s. Given such differences the time at which a sufficient transglottal pressure difference to initiate voicing is reached will vary because the intra-oral-pressure behind the stop drops at different rates.

It is less clear that this factor will explain the parallel differences in the voice onset time in prevocalic voiceless aspirated stops, where the onset of voicing is considerably delayed after the release and any initial differences in flow rate will have dissipated by then. It would be even less obvious that this factor is explanatory if a similar pattern was found in ejective stops, as in this case the pressure built up behind the oral closure will be largely dissipated during the interval between the oral release and the glottal release. Stevens suggests that in aspirated stops a difference in the adduction rate of the vocal folds must be assumed to account for the difference between bilabial and velar aspiration duration, but he does not discuss ejective stops.

In a review of the standard literature and data from a set of 18 less well-known languages Cho and Ladefoged [1] find very general support for the universality of the tendency for bilabial VOT’s to be longer than velar ones in both unaspirated and aspirated voiceless plosives. Coronals are often intermediate, but given the wide range of possible coronal articulations, it is not surprising to find that there is variation in where they fall. Ejective stops occur in six of the languages studied. For this class of stops, Cho and Ladefoged are more impressed with the irregularity of the pattern they observe. However, only two of the languages include bilabial ejectives, and moreover it appears that the measurements examined in this part of their discussion were not all taken in the same manner.

In this paper we will report on further measures of voice onset time in ejective stops and suggest that the VOT pattern is more consistent with that seen in plosives than might have appeared. A possible motivation for the sharing of this pattern across different stop manners will be suggested.

2. Measuring VOT in ejectives
In plosives the standard measure of VOT is from the initial transient of the stop’s oral release to the first voiced pulse of the vowel. A similar measurement can be made on ejectives. However, in an ejective stop there is often a third time point — the release of the glottal closure — which can be identified, as illustrated in Figure 1, which shows an alveolar ejective occurring in a word from the Dagestani language Bagwalal.

![Figure 1. Time measurement points on prevocalic ejective stops. (Spectrogram of the beginning of the word /t'ara/ “run” in Bagwalal).](image-url)

In some studies of ejective timing the interval from the oral release to the glottal release of has been measured as the
VOT, rather than the interval from the oral release to the actual voice onset. In some ejective tokens the two later time points coincide, so that the difference in choice of measurement strategy has no impact on results, but the two methods are measuring different aspects of ejective production and cannot be directly compared.

In Cho and Ladefoged’s comparison the data for three of the six languages with ejectives represent measurements of the ‘standard’ VOT. These are Tlingit, Hupa and Yapese. In the Hupa study [8] the number of tokens measured is very few and the degree of variability high, and these data should probably be treated with some caution. For Navajo [5] and Montana Salish [6] the interval measured is from stop release to glottal release. For the sixth language, Western Apache [7], the interval measured is not explicitly defined, but it seems likely to be the same as was used for Navajo.

Although all six languages have coronal and velar (in the case of Hupa, palatalized velar) ejective stops, only Montana Salish and Yapese have bilabials. Hence there is limited opportunity to compare the VOT of ejective stops at bilabial and velar places and to match this with the most robust place-dependent VOT difference in plosives. Moreover, as noted above, the measurements made in these two languages may not be fully comparable and thus it is possible that the failure of the Montana Salish bilabial/velar comparison to reach significance, in contrast to Yapese, might be due to choice of measurement procedure.

A more extensive set of measurements on Yapese ejectives can now be presented and data from Nez Perce added to extend the comparison.

3. Yapese

Yapese is an Austronesian language spoken on the closely-clustered island group known as Yap in the western Pacific. It is unusual among Austronesian languages in having an extensive set of glottalized consonants including the ejectives /p’, t’, k’/. Measurements on a limited set of tokens elicited in isolation in one-on-one sessions with three speakers by the present author gave mean durations of 60ms for bilabials, 64 for alveolars and 78 for velars, the figures cited by Cho and Ladefoged.

In Figure 2 measurements for tokens elicited in much more casual group recordings with a total of 12 speakers are given. A one-way analysis of variance shows a significant effect of place (F (2,138) = 25.0, p < .0001). Both the bilabial and alveolar VOT’s are significantly shorter than the velar (p < .0001) by Scheffé’s test for comparison of means. Bilabial and alveolar means are not significantly different.

This data conforms to the cross-linguistic trend for velar VOT to be longer than bilabial. Plosive VOT’s have not yet been measured for this extended data set, but in the smaller set measured earlier the mean bilabial plosive VOT was 20 ms, alveolar 22 ms and velar 56, reflecting a very similar pattern to that seen in the ejective stops.

4. Nez Perce

Nez Perce (Niimipuu t’imt) is a Native American language in the Sahaptian group which was formerly spoken by a major community in parts of Idaho, Washington and Oregon. The community was severely disrupted by war in the late 19th century and the language at the moment has few speakers. Lists of isolated words were recorded from five speakers living in or near Lapwai on the Nez Perce reservation in Idaho.

The data set obtained is unfortunately unbalanced for all relevant factors, including position in the word, following vowel and stress contexts, and speaker. Considerable fluctuation was noted with respect to which words are pronounced with ejectives and not all the words sought were known to all speakers. Thus, for example, there are no tokens of /k’, q’/ for speaker 3. The number of tokens of each type obtained from each speaker is shown in Table 1. The analysis that follows is the best that can be done with the materials available.

Table 1: Tokens measured for Nez Perce stop VOT, by speaker and stop type

<table>
<thead>
<tr>
<th>Speaker</th>
<th>Segment</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>Row total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nez Perce</td>
<td>p</td>
<td>15</td>
<td>14</td>
<td>15</td>
<td>17</td>
<td>19</td>
<td>80</td>
</tr>
<tr>
<td>Nez Perce</td>
<td>t</td>
<td>15</td>
<td>18</td>
<td>14</td>
<td>15</td>
<td>17</td>
<td>72</td>
</tr>
<tr>
<td>Nez Perce</td>
<td>k</td>
<td>3</td>
<td>3</td>
<td>11</td>
<td>3</td>
<td>9</td>
<td>29</td>
</tr>
<tr>
<td>Nez Perce</td>
<td>q</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>15</td>
<td>17</td>
<td>87</td>
</tr>
<tr>
<td>Nez Perce</td>
<td>p’</td>
<td>3</td>
<td>5</td>
<td>5</td>
<td>3</td>
<td>9</td>
<td>25</td>
</tr>
<tr>
<td>Nez Perce</td>
<td>t’</td>
<td>15</td>
<td>15</td>
<td>6</td>
<td>12</td>
<td>12</td>
<td>60</td>
</tr>
<tr>
<td>Nez Perce</td>
<td>k’</td>
<td>9</td>
<td>9</td>
<td>0</td>
<td>9</td>
<td>10</td>
<td>37</td>
</tr>
<tr>
<td>Nez Perce</td>
<td>q’</td>
<td>18</td>
<td>12</td>
<td>0</td>
<td>15</td>
<td>17</td>
<td>62</td>
</tr>
<tr>
<td>Nez Perce</td>
<td>Column total</td>
<td>96</td>
<td>94</td>
<td>71</td>
<td>90</td>
<td>108</td>
<td>459</td>
</tr>
</tbody>
</table>

As is shown in Figure 3, the mean durations of plosive and ejective VOT follow very much the same relative pattern of place-dependence except that uvular ejectives have a shorter rather than longer VOT than is found in their velar counterparts. Among plosives, the universal trend is strongly supported in an analysis of variance. There is a significant main effect of place, F (3,271) = 72.7, p < .0001. In post-hoc comparisons of means, /p/ and /h/ are not significantly different, and the /k, q/ comparison is significant only at the .025 level, but all remaining pairwise comparisons are at the p < .0001 level. Among the ejectives the parallel trend is apparent but there is no significant main effect of place (F (3,180) = 1.224, p = .3023), despite the relatively large size of the differences between the means for different places. This

Figure 2. Mean VOT in ms for ejective stops in Yapese. (error bars show one standard error).
is probably largely due to the particularly great variance of the bilabial ejectives. No main effect of place is obtained even if Speaker 3 is omitted and Speaker is added as a separate factor. However, the probability is high that a more controlled data set would show that the trend noted here is reliable.

![Figure 3. Mean VOT in ms for ejective stops in Nez Perce. (error bars show one standard error).](image)

5. Discussion

As we have shown, in two language with bilabial ejectives there is reason to believe that a similar relationship of bilabial to velar voice onset time exists as is found in plosives, the velars having a longer VOT than the bilabials.

In considering this finding, it seems unlikely that the result is to be explained as due to differences in the rate of the articulator movements in the mouth. We offer two hypotheses for consideration. One possibility would be that if place-sensitive differences in unaspirated plosive VOT’s are a natural concomitant of differences in release speed, listeners may come to associate such differences with the place of articulation and generalize them to stops produced with quite different laryngeal gestures.

However, it has often been noted [9] that the relative closure durations of voiceless stops at different places are inversely related to the duration of the VOT which follows them, and this holds for both unaspirated and aspirated plosives. The effect of longer VOT after a shorter closure may thus be primarily a reflection of different closure durations overlaid on a relatively fixed (at least not place-dependent) duration of the glottal gesture which accompanies the stops of different classes. Since much of the data which has been examined in this paper and in Cho and Ladefoged contains primarily utterance-initial stops, the data on closure duration is only sparsely available. Further research widening the durational data to contexts in which closure duration can be measured across a range of stop types is thus a priority to see if the patterns of closure duration observed in plosives also generalize to ejective stops.

We also note that such comparisons can only be made if collection of data from lesser-known languages is encouraged. Even among those languages with an ejective stop series there are relatively few with bilabial ejectives. This occurs in part because some of these languages are generally lacking in bilabial consonants, as in the case of Tlingit. In other cases, there is a specific gap in the ejective series at the bilabial place when bilabial stops in other series do occur. This is the case for Bagwalal, the language exemplified in Figure 1. In this language the only bilabial ejectives occur in a few essentially onomatopoeic words, such as /p’a’pal’a/ “fart”, and their usage is inconsistent across speakers. The reason for this gap would make another paper, but a connection with the tendency for voiceless bilabial stops to be more likely to be absent or rare in phonological inventories is to be suspected.

6. Acknowledgements

The research reported in this paper is supported by a grant from the National Science Foundation for research on the phonetic structures of endangered languages. The help of Julie Larson at UC Berkeley and Heriberto Avelino at UCLA with some of the measurements is gratefully acknowledged.

7. References
