Individual differences in the production and perception of consonant length: implications for sound change

Mary Stevens and Ulrich Reubold (Institute for Phonetics and Speech Processing, Munich)

This study examines a sound change affecting the category boundary between singleton and geminate consonants in Italian e.g. *fato* ‘fate’ v. *fatto* ‘done’. Gemination in Italian involves increased consonant duration and decreased vowel duration and there is evidence that the distinction between the two categories remains robust even across different speaking rates due to the additional language specific temporal cue of the C/V ratio (i.e. C/V > 1 = /C:/ and C/V < 1 = /C/, where C = closure duration and V = preceding vowel; Pickett et al. 1999). While voiceless stops are typically described as unaspirated in Italian, acoustic analysis shows /p: t: k:/ can have aspiration on either side of the closure portion i.e. pre-aspiration [faʰtːo], post-aspiration [faːtʰo] or both [faːtʰo] (e.g. Stevens & Hajek 2010). In previous work Stevens suggests that aspiration allows maintenance of voicelessness with minimal supralaryngeal effort for the speaker and that the overall consonant duration (i.e. [ʰtː] ~ [ʰtː] ~ [tː] ~[tː]) may cue consonant length in both Sienese (Stevens 2012) and in standard Italian (Stevens 2011), rather than closure duration alone. Nonetheless the question arises as to whether native Italian listeners actually assign aspiration to the consonant.

Given pre-aspiration is hard to hear and can alternate with vowel lengthening (e.g. Silverman 2003), Italian listeners may be more likely to assign it to the preceding vowel than the consonant, perceiving e.g. [faʰtːo] as singleton *fato* rather than geminate *fatto*. However Wretling et al. (2003) show that Swedish listeners perceive pre-aspiration as either vowel or consonant lengthening depending on their native dialect. The impact of post-aspiration on consonant length has not been experimentally tested for Italian and it is difficult to predict whether listeners would interpret post-aspirated stops as belonging to either the singleton or the geminate category. The descriptive phonetic literature on regional varieties contains occasional reports of post-aspiration (e.g. Rohlfs 1966, Mele 2009), but it seems that both singletons and geminates can be affected and in any case post-aspiration is not considered a phonetic cue to consonant length (see especially Mele 2009:153 on this point).

The present study tests the impact of aspiration on the perception of consonant length in the variety of Italian spoken in Prato, Tuscany. We also consider the role of inter-listener variation and compare perception results with production data for the same speaker-listeners. In doing so, we attempt to shed new light on historical degemination in Romance: centro-southern varieties including standard Italian are exceptional in having maintained geminates from late Latin e.g. UACCA(M) > Italian /vak:a/ ‘cow’ whereas they were lost almost everywhere else e.g. Spanish /baka/, French /vaf/ ‘cow’. We consider the possibility that historical degemination may have involved pre-aspiration as an intermediate step e.g. [vak:a] > [vaʰk:a] > [vaka].

For the production experiment, speakers read a list of real words containing /p t k p: t: k/.

The duration of the preceding vowel and the voiceless stop (comprising pre-aspiration, closure, post-aspiration) was labeled semi-automatically and the C/V ratio was calculated for each word token. The frequency of pre-aspiration was recorded for each speaker.
For the two alternative forced choice perception experiment we synthesized four *fatto* tokens by manipulating voice termination time for pre-aspiration (VTT) or voice onset time (VOT), keeping vowel and overall consonant (VTT + closure + VOT) duration stable. For the plain unaspirated version, we chose a VOT of 8 ms (*fatto* token 1), and for the post-aspirated version we chose a VOT value of 28 ms (*fatto* token 3). In two additional *fatto* tokens we introduced preaspiration in two different degrees by varying voice termination time (VTT). Subsequent analysis of data from the CLIPS corpus of spoken Italian, as used in e.g. Stevens & Hajek (2010) and Stevens (2011), revealed an inverse correlation of VTT and VOT in preaspirated realisations of geminate /tː/, as can be seen in Figure 1.

![Figure 1: Preaspiration duration (in ms) ~ VOT (in ms) as measured for geminate /tː/ tokens produced with preaspiration. Tokens involve the word *gatto* 'cat', included in a word list read by speakers from 15 Italian cities as part of the CLIPS corpus. One token per speaker (n=40).](image)

The averaged VTT/VOT ratio varied from approximately 6/1 to 1/1 in the data in Figure 1, so for the present stimuli we varied VTT from 48 ms for *fatto* token 2 (with a VOT value of 8 ms as in the plain unaspirated token), to 28 ms for *fatto* token 4 (with a VOT value of 28 ms as in the post-aspirated token). For all four *fatto* tokens the vowel duration was stable at 79 ms, the C/V-ratio (remembering that C = VTT + closure + VOT) was 2.1, but closure duration varied depending on VTT and VOT values. For each of the four *fatto* tokens, a 10-step *fato-fatto* continuum was then synthesized by varying the C/V ratio from 0.6 (= *fato*) to 2.1 (= *fatto*), keeping the duration of the aspiration portions (attributed to C) stable within each continuum. Eight repetitions gave 320 stimuli in total. We predict that if listeners assign aspiration to the consonant there should be no significant difference between plain and aspirated continua whereas if listeners assign aspiration to one of the adjacent vowels there will be more *fato* responses.
Combined results for 20 listeners in Figure 2a show that aspiration influences perception of the singleton-geminate category boundary: whereas the 50%-perceptual boundaries in the plain continuum are to be found consistently at approximately 1 (consistent with Pickett et al. 1999), the rightwards shift for the three aspirated continua shows aspirated stops are less likely to be perceived as geminates than their plain counterparts. An RM ANOVA for the 50%-perceptual boundaries conducted with the factor CONTINUUM and with listeners as random factor revealed a significant effect of CONTINUUM (F[3, 57] = 126.8, p < .001) and post-hoc Bonferroni t-tests revealed significant differences between all 6 possible pairs.

Comparison within each individual listener (by means of Generalized Linear models conducted for each listener’s responses to each continuum) showed no significant differences between plain and pre-aspirated or plain and post-aspirated continua for two listeners (GM and EP, see also Figure 3), and one of these (GM) showed no difference between all four continua. All other listeners tended to assign both aspiration types to the relevant adjacent vowel.

Figure 2a) Proportion of geminate responses (averaged on 20 listeners) and 1b) boxplots showing the cross-over point from fato to fatto responses for the four continua.

In terms of production, speaker-specific differences are evident in the frequency of pre-aspiration (consistent with other studies) which ranges from 0-37% of /t:/ tokens analyzed. Preliminary speaker-specific comparisons (n=10) suggest that the production and perception of pre-aspiration are linked inasmuch as speakers with a tendency to produce geminates with pre-aspiration are also more likely to attribute pre-aspiration to the consonant (cf. Figure 3), but not all speakers conformed to this pattern. For example AF and ER in Figure 3 clearly perceive differences between the pre-aspirated and the non-aspirated continuum, and yet produce pre-aspirated stops in approximately 30% of all geminate productions, whereas the aforementioned speaker-listeners GM and EP also produce pre-aspirated geminate stops in approximately 30% of the cases (and therefore do so more than others), but do not distinguish perceptually between pre- and non-aspirated continua.
Figure 3. Percentage of pre-aspirated geminate stops in production plotted against the difference between the pre-aspirated (red in Figure 2a) and non-aspirated (black in Figure 2a) continuum (in units of stimulus number) for 10 speaker-listeners. Cook’s distance, an estimate of the influence of a data point on least squares regression analysis as performed here for creating the regression line and therefore a detection method for outliers, is represented by means of bubble plots (size of red circles corresponds to the value of Cook’s distance).

The role of post-aspiration in the perception data, as well as interactions between the C/V ratio and aspiration in our production data are yet to be addressed. A first indication of how speaker-listeners parse their own productions of whether pre-aspiration is assigned to the consonant or to the vowel in production of geminate stops is to compare each speaker-listener’s CV-ratio in non-pre-aspirated productions (of the words *ditta, facto, gatta*, and *matti*) with two measures of CV-ratio in their corresponding pre-aspirated geminate stops, one assigning the pre-aspiration duration to the consonant, one assigning it to the vowel, i.e.
\[
C/V_{\text{pre_vowel}} = (\text{closure} + \text{VOT}) / (\text{vowel} + \text{VTT})
\]

and

\[
C/V_{\text{pre_consonant}} = (\text{VTT} + \text{closure} + \text{VOT}) / \text{vowel}
\]

Again, speaker-listener ER stands out, as she is the only one whose CV-ratio values for pre-aspirated stops are more similar to those of plain productions when the pre-aspiration duration is assigned to the consonant (cf. final column of Table 1). EB and EC are also different from the other speaker-listeners, because in these two cases it is difficult to know whether to assign pre-aspiration to the consonant or to the vowel.

<table>
<thead>
<tr>
<th>Speaker-listener</th>
<th>CV_ratio (plain)</th>
<th>C/V_{\text{pre_consonant}}</th>
<th>C/V_{\text{pre_vowel}}</th>
<th>Nearer to plain</th>
</tr>
</thead>
<tbody>
<tr>
<td>AF</td>
<td>1.9</td>
<td>3</td>
<td>1.2</td>
<td>V</td>
</tr>
<tr>
<td>AP</td>
<td>1.5</td>
<td>2.1</td>
<td>1.2</td>
<td>V</td>
</tr>
<tr>
<td>CC</td>
<td>1.9</td>
<td>3.8</td>
<td>1.8</td>
<td>V</td>
</tr>
<tr>
<td>EB</td>
<td>1.4</td>
<td>1.9</td>
<td>1</td>
<td>V≈C</td>
</tr>
<tr>
<td>EC</td>
<td>1.4</td>
<td>1.9</td>
<td>1</td>
<td>V≈C</td>
</tr>
<tr>
<td>EP</td>
<td>1.8</td>
<td>3.7</td>
<td>1.4</td>
<td>V</td>
</tr>
<tr>
<td>ER</td>
<td>2.2</td>
<td>2.7</td>
<td>1.4</td>
<td>C</td>
</tr>
<tr>
<td>FS</td>
<td>1.8</td>
<td>3.2</td>
<td>1.6</td>
<td>V</td>
</tr>
<tr>
<td>GM</td>
<td>1.8</td>
<td>2.7</td>
<td>1.3</td>
<td>V</td>
</tr>
</tbody>
</table>

Table 1: Comparison of C/V_{\text{pre_consonant}} and C/V_{\text{pre_vowel}} with CV-ratio values of non-pre-aspirated productions

The general tendency to assign pre-aspiration to the preceding vowel in Figure 2a provides some experimental support for the notion that historical degemination in Romance may have come about via pre-aspiration. However individual differences in the production and the perception of aspiration within our homogenous group show that it is not always clear whether pre-aspiration can be assigned to the vowel or the consonant. This provides some experimental evidence that individual differences in perception and production may influence the directionality of sound changes involving aspiration, giving rise to the dialectal differences noted for pre-aspiration in Swedish, for example.

References


