The consonant length contrast in Persian: Production and perception

Benjamin B. Hansen & Scott Myers
Department of Linguistics, University of Texas at Austin

Abstract
In Persian, length contrasts occur in all classes of consonants, including obstruents, glides and laryngeals. The results of an acoustic study of Persian consonant length are reported, showing that geminates differ from singletons across speaking rates and manner classes in constriction duration, preceding vowel duration, formant transition duration, and the intensity drop from the preceding vowel. Two perception studies tested the claim that consonant length contrasts are more difficult for Persian speakers to perceive when the consonant is more similar to neighboring vowels, as is the case with glides and laryngeals. In one of these experiments, Persian speakers identified the length class of consonants differing in constriction duration and consonant type. Identification was most consistent and rapid when the test consonant was an obstruent, and least so when it was laryngeal. In the other perception experiment, subjects identified the length class in consonants differing in both constriction duration and formant transition duration. Consonants with longer formant transitions were more likely to be identified as long than consonants with shorter transitions, but also were identified less consistently and more slowly. These results suggest a phonetic explanation for the avoidance of length contrasts in more vowel-like consonants in many languages.
The consonant length contrast in Persian: Production and perception

1 Introduction

Persian (Farsi) has a contrast in consonant length, as illustrated by such minimal pairs as fæʔal (verb) – fæʔːal (active thing), or bænɑ (building) – bænːa (builder). This contrast extends to all the consonant classes in the language, listed in Table 1 (Samareh 1977, 1985, Mahootian 1997, Majidi & Ternes 1999, Deyhime 2000, Bijankhan & Nourbakhsh 2009, Hansen 2004, 2012).

Table 1. Consonants of Persian

<table>
<thead>
<tr>
<th></th>
<th>Bilabial</th>
<th>Labiodental</th>
<th>Dental/Alveolar</th>
<th>Post-alveolar</th>
<th>Palatal</th>
<th>Uvular</th>
<th>Glottal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Voiceless obstruent stop</td>
<td>p</td>
<td>t</td>
<td>tʃ</td>
<td>c</td>
<td>q</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Voiced obstruent stop</td>
<td>b</td>
<td>d</td>
<td>dʒ</td>
<td>j</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Voiceless fricative</td>
<td>f</td>
<td>s</td>
<td>ʃ</td>
<td>x</td>
<td>h</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Voiced fricative</td>
<td>v</td>
<td>z</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nasal stop</td>
<td>m</td>
<td>n</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alveolar trill</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lateral approximant</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Approximant glide</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>j</td>
<td></td>
</tr>
</tbody>
</table>
The range of the length contrast in Persian is striking because in many languages the most vowel-like consonants, such as glides, liquids, and laryngeals, are excluded from the contrast (Podesva 2000, Kawahara 2007, Maddieson 2008, Kawahara 2012, Kawahara, Pangilinan & Garvey submitted). For example, Icelandic (Garnes 1969) and Classical Nahuatl (Andrews 1975) have length contrasts for all classes of consonants except for glides. Koya (Tyler 1969: 34) and Telugu (Kostić, Mitter & Krishnamurti 1977) have length contrasts in all consonants except for laryngeals. Length contrasts in Luganda (Ashton et al. 1954) include all consonants except for glides and liquids, and those in Toba Batak (Nababan 1981) and Finnish (Lehtonen 1970) include all consonants except for glides and laryngeals. Glides, liquids and laryngeals are excluded from length contrasts in Punjabi (Malik 1995) and Pengo (Burrow & Bhattacharya 1970). Other languages combine such restrictions with other restrictions on what classes of consonants can be long, e.g. Japanese (no length contrasts in glides, liquids, laryngeals, or voiced obstruents: Kawahara 2007), or Tamazight Berber (no length contrasts in glides or voiced obstruents: Penchoen 1973).

Podesva (2000) and Kawahara (2007) express these cross-linguistic tendencies with Optimality-Theoretic markedness constraints which penalize long consonants (geminates) with higher sonority. Both authors suggest that these phonological constraints have a phonetic basis in the difficulty of fixing the acoustic boundary between a sonorant consonant and a neighboring vowel (the canonical position for long consonants being between two vowels). The transition between such a consonant and a vowel tends to be gradual, so that it is difficult to determine where the first sound in the sequence ends and the second begins (Peterson and Lehiste 1960, Myers & Hansen 2005).
Kawahara (2007) demonstrated this perceptual challenge in an experiment with native speakers of Egyptian Arabic, a language in which the length contrast extends to all consonant classes. Participants heard stimuli in which the duration of the constriction interval of an intervocalic consonant was varied step-wise, and they had to identify the medial consonant as long or short. Kawahara found that the participants’ reaction times were significantly longer when the test consonant was a glide than when it was a liquid, and longer for sonorant consonants than for obstruent ones. He also found that the slope of the transition region of the identification curve was steeper for obstruents than for sonorants, and for liquids compared to glides. He concluded that the length contrast in Egyptian Arabic is less perceptible in glides than in liquids, and less perceptible in liquids than in nasals or obstruents.

In subsequent work, Kawahara has explored the bases of these perceptibility differences in non-speech analogues. Kawahara (2012) found that in sine waves with a medial low-intensity interval (analogous to a medial consonant), discrimination and identification of differences in the duration of that interval were more accurate when the difference in intensity between medial low-intensity interval and the surrounding high-intensity intervals was greater. Kawahara, Pangilinan, and Garvey (submitted) synthesized sine wave complexes with 50 sine wave components, and a medial interval of silence, noise, or a lower-amplitude sine wave complex with the same spectral specifications as the surrounding intervals. They found that differences in the duration of the medial “consonant” interval were more accurately detected where that interval consisted of silence or noise than when it consisted of the low-amplitude periodic interval. The two sets of experiments taken together provide evidence that length contrasts are harder to perceive the more acoustically similar the contrasting element is to the surrounding elements.
Such perceptual difficulties could lead to the avoidance of length contrasts in more vowel-like consonant classes. If a contrast is difficult to perceive, that increases the likelihood that a language learner will fail to master it. Such a language learner would then lack the contrast, and if enough members of a speech community follow that learner, the contrast would be lost in the language. The tendency to misperceive a sound or sound sequence in a particular way is the basis for sound change and for recurring gaps in segment inventories (Ohala 1981, 1983, 1993, Wedel 2007).

The widespread absence of length contrasts in glides and laryngeals provides perspective for the contrast between geminate and singleton consonants in Persian. Given the long gradual transitions between a vowel and a neighboring glide, we might expect greater difficulty for Persian speakers in distinguishing contrasting length classes in glides compared to other consonant types, such as nasals, lateral approximants, or obstruents. Likewise, similar difficulties would be expected in the length contrast in the laryngeals [h] and [ʔ], since there is no formant transition between the laryngeal and a neighboring vowel (Lehiste 1964, Keating 1988). Reduced realizations of [h] and [ʔ] frequently consist of an interval of breathy- or creaky-voicing (respectively) in the vowel formants (Pierrehumbert and Talkin 1992). It is challenging to determine a boundary between such a reduced laryngeal and a neighboring vowel, just as it is with a glide.

These segmentation issues are illustrated in Fig. 1 in three spectrograms of Persian [æCɑ] sequences. The steady state interval of the medial consonant is marked off by blue boundaries, with the midpoint marked in red.
Figure (1a) with a medial palatal glide [j] displays the gradual formant transitions from preceding vowel to the steady state interval of the glide and from there to the following vowel.

Figure (1b) with medial [h] shows the diphthongal formant transition from the preceding to
following vowel proceeding unimpeded through the voiceless interval that constitutes the consonant. The gradual transition between modal voicing in the vowels and voicelessness in the [h] can be seen in the not fully modal pulses just outside the marked voiceless interval. Figure (1c) with a medial [ʔ] shows a typical realization of that category as an interval of creaky-voicing. As in the [h], the formant values of the neighboring vowels continue through the creaky-voiced interval that constitutes the consonant, and the gradualness of the voice quality transition can again be seen in the weakness of the pulses neighboring the consonant interval.

We hypothesize that discrimination of Persian consonants contrasting in length is more difficult the more vowel-like the consonant is. The more acoustically similar a consonant is to the neighboring vowel, the more indeterminate the boundary between the two segments, and thus the more indeterminate the duration of that consonant. Glides are similar to vowels in having prominent formants and in being relatively intense, and the transition between a glide and a neighboring vowel is long. Laryngeals are similar to neighboring vowels in sharing the same formant values, while differing from them in intensity and voice quality. With other classes of consonants, such as nasals and obstruents, the acoustic boundary between consonant and neighboring vowel is more distinct, which eases the perception of length contrasts in those consonants.

We first consider how the length contrast is acoustically realized across consonant classes and speaking rates in Persian, presenting results from a production study. We then present the results of two perception studies investigating how identification of consonant length classes varies across consonant classes, and how the length of the vowel-consonant transition affects that identification.
Experiment #1: Acoustic differences between singleton and geminate consonants across speaking rates and consonant classes

A production experiment was run to determine how the acoustic differences between long and short consonants varied according to consonant class in Persian. Medial consonants were included belonging to the classes of obstruents, nasals, liquids, glides, and laryngeals.

Like other duration measures, consonant duration is shorter at faster speaking rates than at slower ones (Pickett, Blumstein & Burton 1999, Arvaniti 1999, Hirata & Whiton 2005). Hansen (2004) found that the consonant length contrast in Persian is maintained across speaking rates, but the difference between singleton and geminate consonants is reduced at faster rates, since geminate consonants vary more due to rate than singleton consonants. Due to this strong interaction of the length contrast with speaking rate, variation in speaking rate was included in the production experiment, to test whether the differences between geminates and singletons hold across different rate conditions.

2.1 Methods

2.1.1 Participants

Four adult native speakers of Persian, currently residing in the U.S., participated in the study. Two were female (F1, F2) and two male (M1, M2). Their ages ranged from 38 to 44 at the time of the study. All participants grew up in Tehran, stayed there at least through secondary
school, and continue to speak the language on a daily basis. All four are fluent speakers of English as well.

2.1.2 Materials

The test consonants were the long and short counterparts of the consonants [h, ʔ, j, l, n, z, d]. Each consonant occurred in medial position in a Persian word between a preceding low front vowel [æ] and a following low back vowel [ɑ], in a syllable bearing main word stress. The 14 test words are listed in Table 2, with the test consonant boldfaced.

<table>
<thead>
<tr>
<th>Segment type</th>
<th>Singleton</th>
<th>Geminate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Voiceless laryngeal continuant</td>
<td>dʒæhan (world)</td>
<td>sæhaf (book-binder)</td>
</tr>
<tr>
<td>Voiceless glottal stop</td>
<td>fæʔal (verb)</td>
<td>fæʔal (active thing)</td>
</tr>
<tr>
<td>Voiced palatal glide</td>
<td>bæjan (exposition)</td>
<td>xæjan (surname)</td>
</tr>
<tr>
<td>Voiced alveolar lateral approximant</td>
<td>bæla (haughty one)</td>
<td>dælal (laborer)</td>
</tr>
<tr>
<td>Voiced alveolar nasal stop</td>
<td>bæna (building)</td>
<td>bæna (builder)</td>
</tr>
<tr>
<td>Voiced alveolar fricative</td>
<td>qæza (food)</td>
<td>bæzəaz (rug merchant)</td>
</tr>
<tr>
<td>Voiced dental plosive</td>
<td>fæḍa (sacrifice)</td>
<td>mæḍτah (eulogist)</td>
</tr>
</tbody>
</table>

The test words were produced in the constant carrier frame Minu ___-ra did (“Minu saw ___”).
2.1.3 Procedure

The sentences were presented to the subjects in Persian script on sheets of paper, with filler items at the beginning and end of each page. There were 12 sheets, each with all 14 sentences in a different randomized order. Participants were asked to read the whole set of 168 sentences (12 sheets * 14 sentences) without sentence-internal pauses, at a normal, relaxed rate. Then they were asked to read the same list of items again at a faster rate, though still speaking clearly enough to be understood. Finally, they were asked to read the sentences a third time speaking as fast as possible, without worrying about whether they were understood. Each participant thus produced 168 sentences in three rate conditions (baseline/faster/fastest), for a total of 504 sentences per participant. The readings were recorded in a sound-treated recording booth on a solid-state digital recorder at a sampling rate of 22,000 Hz at 16-bit amplitude precision.

2.1.4 Measurements and hypotheses

The chief acoustic and perceptual correlate of a consonantal length contrast is the duration of the constriction interval (Obrecht 1965, Lehtonen 1970, Lahiri & Hankamer 1988, Aoyama & Reid 2006). This was measured in a waveform display, with reference to a simultaneous spectrogram display, using Kay Elemetrics MultiSpeech. The onset of the consonant was at the end of the decline in intensity from the vowel, at the point at which the waveform attained its local minimum of intensity and complexity. The offset of the consonant was the end of that interval of low intensity and wave complexity. In the glide [j], where there
often was no such clear low-point in intensity, the onset of the glide was placed at the point at which F2 reached its local maximum and F1 its local minimum (whichever event took place last), and the offset was the end of the interval with those F1 and F2 values. It was expected that the constriction would be longer for geminates than for singletons, and longer at slower rates than at faster ones.

The duration of the vowel preceding the test consonants was also measured. This was measured from the onset of voicing after the release of the preceding consonant, to the onset of the test consonant. In many languages, vowels are systematically shorter before geminate consonants than before singletons (Maddieson 1985). But Hansen (2004) found that vowels in Persian were on the contrary significantly longer before geminates than before singleton consonants, an effect also found in Finnish (Lehtonen 1970) and Japanese (Han 1994, Smith 1995). Smith (1995) suggests that these cross-linguistic differences in the effects of consonant length on the duration of a preceding vowel reflect differences between languages in the pattern of intergestural timing of vowels and consonants.

The duration of the formant transition between the preceding vowel and the test consonant was measured. The onset of this interval was the onset of the movement in F1 or F2 (whichever changed first) from the values for the preceding vowel towards those for the test consonant. It was expected that geminates would have longer formant transitions than singletons, reflecting a lower velocity of articulator movement (and lower gestural stiffness) in producing the geminate, as found in articulatory studies by Smith (1995) and Löfqvist (2007) for Japanese. Myers and Hansen (2005) found that long vowels have longer formant transitions than short vowels in Finnish. The duration of the formant transition is also important in that longer formant transitions blur the boundary between the successive segments, the transition being the interval
in which the acoustic output is determined in varying degrees by each of two successive segments (Öhman 1967, Myers and Hansen 2005). No formant transition duration could be measured for the laryngeals [ʔ, h] since they share the formant values of the neighboring vowels, so that any formant movement in the neighbourhood of the laryngeal was a diphthongal one between the two successive vowels (Lehiste 1964, Keating 1988).

The difference in intensity (dB) was measured between the midpoint of the preceding vowel and the midpoint of the constriction interval of the test consonant. It was expected that the drop in intensity from vowel to following consonant would depend on the degree of stricture of the consonant, with more constricted articulations associated with a greater drop in intensity (Byrd & Tan 1996, Van Son & Pols 1999, Mauk 2003, Warner & Tucker 2011). This gives one measure of acoustic similarity of the consonant and neighboring vowel. It was also expected that this drop in intensity would be greater in geminates than in singletons, due to greater coarticulatory reduction of the consonant constriction under time pressure when the constriction is briefer.

As a measure of rate, the duration of the frame was measured. This was the duration of the whole utterance, minus the duration of the test word. Removing the duration of the test word left only the duration of the constant frame Minu ___-ra did. It was expected that this duration would be greater for slower speaking rates than for faster ones. Its importance in the present study is that it allows us to factor out the effects of rate, and present the effects of consonant length and consonant type across speaking rates.
2.2 Results

The results will be presented in the following sections by measurement. All statistical tests were conducted with a mixed-model regression analysis, using the *lme4* package in *R*, with the participant as a random factor. The analyses were hierarchical (Cohen and Cohen 1975), with the frame duration as the first independent variable. The residuals of that analysis were then treated as the dependent variable in an analysis with the independent variables Length (geminate/singleton), Obstruent (nonobstruent/obstruent), Glide (nonglide/glide) and Laryngeal (nonlaryngeal/laryngeal). The latter, italicized group in each factor was treated as the marked value (coded as 1), while the other group was treated as a default (coded as 0). P-values were estimated using the Markov Chain Monte Carlo sampling with the *pvals.fnc* function in *R*.

2.2.1 Constriction duration

As expected, the mean constriction duration was longer in geminate (108 ms) than in singleton consonants (56 ms). It was longer at slower rates than at faster ones: normal (103 ms) – faster (84 ms) – fastest (60 ms). As found by Hansen (2004), the effect of rate increases was greater for geminates than for singletons, as seen in Fig 2, which shows the distribution of constriction duration values according to length and rate.
Figure 2. Constriction duration by consonant length class and speaking rate
The mean constriction duration for geminates ranged from 137 ms in the baseline rate to 76 ms at the fastest rate, while the range for singletons was from 69 ms for the baseline rate to 43 ms for the fastest rate. As a result of this asymmetry in range, the mean constriction duration of singletons was 50% that of geminates at the normal and faster rates, but 57% at the fastest rate.

The mean constriction duration values for each segment class and each length class are given in Table 3:

Table 3. Mean constriction duration (ms) by length class and segment class

<table>
<thead>
<tr>
<th></th>
<th>[h]</th>
<th>[ʔ]</th>
<th>[j]</th>
<th>[l]</th>
<th>[n]</th>
<th>[z]</th>
<th>[d]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Geminate</td>
<td>131</td>
<td>116</td>
<td>63</td>
<td>110</td>
<td>113</td>
<td>113</td>
<td>114</td>
</tr>
<tr>
<td>Singleton</td>
<td>75</td>
<td>58</td>
<td>36</td>
<td>55</td>
<td>49</td>
<td>77</td>
<td>44</td>
</tr>
<tr>
<td>Ratio of geminate to singleton means</td>
<td>1.75</td>
<td>2.00</td>
<td>1.75</td>
<td>2.00</td>
<td>2.31</td>
<td>1.47</td>
<td>2.59</td>
</tr>
</tbody>
</table>

In every segment class, the mean constriction duration for geminates was greater than that for the corresponding singletons. The greatest proportional differences between geminate and singleton were in the stops ([n, d̪, ʔ]) and [l], which shares with the stops the fact that it is made with contact of the articulator with the place of articulation. The smallest differences between geminate and singleton were in the continuants [z, j, h].

Rate had a significant main effect on constriction duration (t = 20.2, p < .001), with longer frame durations associated with longer constriction durations. With that effect factored out, there were significant main effects of Length (t = -33.2, p < .001), Glide (t = -20.8, p < .001), and Laryngeal (t = 7.0, p < .001), but not Obstruent (p = .09). There was a significant
interaction between Length and Glide ($t = 9.9$, $p < .001$), and between Length and Obstruent ($t = 2.4$, $p = .02$).

The effects of segment classes Glide, Laryngeal and (in interaction) Obstruent reflect inherent duration differences among consonant classes (Klatt 1976). To explore how these class-dependent duration differences interacted with the effects of consonant length, a separate mixed-model analysis was performed for each consonant class, in which the effects of rate were first factored out, and then the effects of Length on the residuals then tested. To compensate for the multiple comparisons, the alpha value was adjusted by means of the Bonferroni correction to .007 (.05/7). Geminates were significantly longer in all manner classes: [h] ($t = -18.8$, $p < .001$), [ʔ] ($t = -20.4$, $p < .001$), [j] ($t = -22.2$, $p < .001$), [l] ($t = -27.6$, $p < .001$), [n] ($t = -31.9$, $p < .001$), [z] ($t = -19.0$, $p < .001$), and [d̪] ($t = -33.0$, $p < .001$).

It was noted above that the continuants [z, j, h] had a lower geminate/singleton ratio than the noncontinuants [n, d̪, ʔ, l]. An analysis was therefore undertaken in which the single factor Continuant (continuant/noncontinuant) replaced the consonant class factors Glide, Laryngeal and Obstruent. After factoring out the effect of rate, it was found that both Length and Continuant had significant main effects: Length ($t = -23.4$, $p < .001$), Continuant ($t = 5.3$, $p < .001$).

Moreover, the interaction of the two factors was significant ($t = -8.7$, $p < .001$), reflecting the fact that the difference between geminate and singletons was greater in the noncontinuants.

Geminates had a longer constriction duration than singletons, across speaking rates and consonant classes. The greatest differences between geminate and singleton lay in those consonants that were produced with contact, i.e. the stops and lateral approximant [l]. Similarly,
Aoyama & Reid (2006) found in Guinang Bontok that the smallest geminate/singleton duration ratios in that language belonged to the continuants /j, s, ɹ, w/, and the greatest ratios belonged to the stops and the lateral approximant: /ŋ, p, k, t, m, n, l/.

### 2.2.2 Preceding vowel duration

The mean duration of a vowel preceding a geminate (105 ms) was slightly longer than that of a vowel preceding a singleton (99 ms). The mean duration of this vowel was shorter at faster rates than at slower ones: normal (118 ms) vs. faster (103 ms) vs. fastest (86 ms). As with constriction duration, the difference in preceding vowel duration between geminate and singletons is less at faster rates than at slower rates. The proportion of the mean preceding vowel duration for singletons compared to geminates was .90 for the normal rate, .94 for faster, and .99 for the fastest.

The effect of consonant length on the duration of the preceding vowel varied considerably according to the segment type of the consonant, as can be seen in Table 4.

<table>
<thead>
<tr>
<th></th>
<th>[h]</th>
<th>[ʔ]</th>
<th>[j]</th>
<th>[l]</th>
<th>[n]</th>
<th>[z]</th>
<th>[d]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Geminate</td>
<td>82.2</td>
<td>63.9</td>
<td>116.2</td>
<td>123.9</td>
<td>116.7</td>
<td>133.3</td>
<td>101.3</td>
</tr>
<tr>
<td>Singleton</td>
<td>93.1</td>
<td>61.9</td>
<td>115.6</td>
<td>106.4</td>
<td>113.6</td>
<td>105.4</td>
<td>96.4</td>
</tr>
<tr>
<td>Ratio of geminate to singleton means</td>
<td>0.88</td>
<td>1.03</td>
<td>1.01</td>
<td>1.16</td>
<td>1.03</td>
<td>1.26</td>
<td>1.05</td>
</tr>
</tbody>
</table>
The vowel was longer before a geminate than before a singleton, except with the laryngeal [h], with which the preceding vowel was 12% shorter before the geminate. The greatest differences between geminate and singleton in preceding vowel duration were with [z] (with a 26% difference) and [l] (with a 16% difference).

Vowel duration was significantly affected by speaking rate, with greater vowel duration associated with greater frame duration (t = 26.0, p < .001). Factoring out that effect, there were main effects of Length (t = -7.9, p < .001) and Laryngeal (t = -33.0, p < .001), but not of Glide (p = .15) or Obstruent (p = .08). All interactions were significant: Length * Obstruent (t = -2.9, p = .003), Length * Glide (t = 3.6, p < .001), Length * Laryngeal (t = 2.0, p < .001).

To explore the interaction of segment type with length, separate tests were performed for each consonant class as before with constriction duration. The vowel was significantly longer before geminates than before singletons when the consonant was [l] (t = -12.8, p < .001), [n] (t = -3.7, p < .001), [z] (t = -21.4, p < .001), or [d̪] (t = -4.6, p < .001). When the consonant was [h], the vowel was significantly shorter before a geminate than before a singleton (t = 4.6, p < .001). With [ʔ] or [j], the effect of consonant length on preceding vowel duration was not significant at the adjusted .007 level.

Thus, contrary to the finding of Hansen (2004), it was not found in this study that the vowel was consistently longer before geminates than before singletons. The earlier study was limited to the dental stops [d̪] and [t̪], and the finding was replicated for those consonants, but it did not hold for the laryngeals [h] and [ʔ], or for the glide [j]. This suggests that the durational
effect of consonant length on preceding vowel duration in Persian is not a general effect of interarticular timing sensitive to syllable position, as suggested for Japanese by Smith (1995).

2.2.3 Formant transition duration

The mean formant transition duration was longer from a vowel to a geminate (47 ms) than from a vowel to a singleton (42 ms). As with other duration measures, it was shorter with faster rates: 50 ms in baseline rate, 46 ms at the faster rate, and 38 ms at the fastest rate.

Table 5 gives the mean formant transition duration broken down by length class and consonant class. The laryngeals [ʔ, h] are excluded since they had no formant transitions from the preceding vowel.

**Table 5.** Mean formant transition duration by length and consonant class

<table>
<thead>
<tr>
<th></th>
<th>[j]</th>
<th>[l]</th>
<th>[n]</th>
<th>[z]</th>
<th>[d]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Geminate</td>
<td>86.0</td>
<td>32.3</td>
<td>34.4</td>
<td>56.7</td>
<td>39.2</td>
</tr>
<tr>
<td>Singleton</td>
<td>71.6</td>
<td>33.0</td>
<td>32.3</td>
<td>49.5</td>
<td>38.4</td>
</tr>
<tr>
<td>Ratio of geminate to singleton means</td>
<td>1.20</td>
<td>0.98</td>
<td>1.07</td>
<td>1.14</td>
<td>1.02</td>
</tr>
</tbody>
</table>

The mean formant transition was longer in geminates than in singletons for every class except [l]. It was longer in the continuants than in the noncontinuants, and the geminate/singleton ratio was greater.

The effect of rate on transition duration was significant, with longer frame duration associated with longer transition duration (t = 7.9, p < .001). With this effect factored out, there
were significant main effects of Glide (t = 43.8, p < .001) and Obstruent (t = 14.9, p < .001), but not of Length (p = .31). There were, however, significant interactions of Length with Glide (t = -8.13, p < .001) and with Obstruent (t = -2.3, p = .02), reflecting the differences in the geminate-singleton ratio across the consonant classes.

The interactions were explored by testing the effect of Length in each consonant class. The transition duration was significantly longer in geminates than in singletons in continuants [j] (t = -11.0, p < .001) and [z] (t = -6.9, p < .001), but not in noncontinuants [n] (p = .04), [d] (p = .33), or [l] (p = .57).

With Continuant as the predictor variable and the effects of rate factored out, there were significant main effects of both Length (t = -9.8, p < .001) and Continuant (t = -34.6, p < .001), and a significant interaction between the two (t = 6.9, p < .001). The interaction reflects the greater difference between singletons and geminates in continuants compared to noncontinuants.

Thus the consonant length contrast was reflected in formant transition duration, but only for the continuants. These were the consonant types with the smallest geminate-singleton ratios in constriction duration in Table 2, so it could be that there is a trading relation between the two measures, with bigger consonant length differences in one measure allowing smaller differences in the other.

By a considerable margin, the longest formant transition duration values were for the glide [j], with a transition duration of 86.0 ms for geminates and 76.1 ms for singletons. If it is the case that longer formant transitions blur the boundaries between segments and make segmentation more difficult, then the glide presents particular challenges for length classification.
2.2.4 Intensity difference from preceding vowel

The mean intensity drop from vowel to following consonant was greater for geminates (12.1 dB) than for singletons (6.3 dB). The mean drop was also reduced at faster speaking rates: normal (10.4 dB) compared to faster (9.5 dB) and fastest (7.7 dB).

Table 6 presents the mean intensity drop broken down by length and consonant class.

**Table 6.** Mean intensity difference between vowel and following consonant by length and consonant type

<table>
<thead>
<tr>
<th></th>
<th>[h]</th>
<th>[ʔ]</th>
<th>[j]</th>
<th>[l]</th>
<th>[n]</th>
<th>[z]</th>
<th>[d]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Geminate</td>
<td>15.9</td>
<td>27.3</td>
<td>3.9</td>
<td>2.3</td>
<td>2.5</td>
<td>11.5</td>
<td>21.1</td>
</tr>
<tr>
<td>Singleton</td>
<td>7.8</td>
<td>9.9</td>
<td>1.9</td>
<td>2.1</td>
<td>2.2</td>
<td>10.5</td>
<td>9.6</td>
</tr>
<tr>
<td>Ratio of geminate to singleton means</td>
<td>2.03</td>
<td>2.75</td>
<td>2.15</td>
<td>1.10</td>
<td>1.12</td>
<td>1.09</td>
<td>2.20</td>
</tr>
</tbody>
</table>

The intensity drop was greater to consonant types with lower inherent intensity, so that the obstruent stop [d̪] had a greater drop than the fricative [z] and both obstruents had a greater drop than the sonorants [j, l, n]. The laryngeals [h, ʔ] class in this regard with the low-intensity obstruents.

The effect of rate on the intensity drop was significant, with longer frame duration associated with greater intensity drop (t = 6.1, p < .001). With this effect factored out, there were main effects of Glide (t = 2.6, p = .01), Laryngeal (t = 35.5, p < .001) and Obstruent (t = 25.7, p < .001), but not Length (p = .55). However, there were significant interactions of Length with
Glide (t = -2.1, p = .04), Laryngeal (t = -16.2, p < .001) and Obstruent (t = -7.8, p < .001), reflecting the fact that the geminate/singleton difference varies by consonant type.

Breaking down the data into consonant subsets, geminates were found to have a significantly greater intensity drop than in singletons in [h] (t = -14.3, p < .001), [ʔ] (t = -14.4, p = .001), [j] (t = -7.4, p < .001), and [d̪] (t = -16.4, p < .001), but not [l] (p = .44), [n] (p = .28) or [z] (p = .02).

The intensity drop from vowel to consonant was greater in geminates than in singletons, and greater at slower rates. This presumably reflects the greater degree of undershoot and reduction where the constriction is briefer (Byrd & Tan 1996, van Son & Pols 1999, Mauk 2003, Warner & Tucker 2011). In the consonant classes, the greatest intensity drops were for the lowest intensity consonants, the laryngeals and obstruents, and the least for the higher intensity sonorant consonants. With respect to intensity, then, it is the sonorant consonants which are most similar to the neighboring vowel, and the obstruents and laryngeals which are least similar.

2.2.5 Summary of production study

The production study has revealed a number of ways in which geminate consonants in Persian differ acoustically from the corresponding singletons. First and foremost, the constriction interval was longer in geminates than in singletons, and this difference held across rate and consonant types.

Geminates were also distinguished from the corresponding singletons in having a longer preceding vowel duration, a longer formant transition duration, and a greater intensity drop, but
these differences held only for some consonant types. Preceding vowel duration could be used to distinguish geminates from singletons except with [ʔ] and [j]. Formant transition distinguishes the length classes only for the continuants, and the intensity drop for all consonants except [l, n, z].

We were also considering the similarity of the consonant to the preceding vowel, with an eye to considering how vowel-likeness affects segmentability and thus the perceptibility of consonant length classes. In terms of intensity, the consonant types with the smallest difference from the preceding vowel were the glides and other sonorant consonants, while the laryngeals and obstruents were the most distinct from the preceding vowel.

On the other hand, in terms of formant transitions, the glides were the consonants with the longest formant transitions, blurring the acoustic boundary between consonant and vowel. The laryngeals had no formant transitions from the previous vowel, so that formant movement is unavailable as a means of determining the boundary between vowel and consonant. Both kinds of transition make it more difficult to determine the acoustic boundary between consonant and vowel, and so potentially make it more difficult to determine the consonant’s length class. The following perception experiments will test the claim that both laryngeals and glides are more difficult to assign to length classes than other classes of consonants.

3 Experiment #2: Effect of consonant type on length class identification

A perception experiment was designed to test how the differences among the consonant types are reflected in the identification of geminate consonants by Persian-speaking listeners.
The overarching hypothesis is that length identification will be slower and more equivocal for more vowel-like segment types (glides and laryngeals).

3.1 Methods

3.1.1 Participants

The participants in this study were 5 female and 5 male adult native speakers of Persian, who had grown up in Tehran and stayed there through at least high school. All participants were residing in the U.S. at the time of the experiment, but continued to use Persian regularly in family interactions. One male and one female participant in this experiment had also participated in the production study. One male participant failed to complete the experiment, so his responses were not included in the analysis.

3.1.2 Stimuli

The stimuli were nonsense words varying in constriction duration of a medial consonant. A female speaker of Tehran Persian produced nonsense words of the form [qæC:ɑb], where C belonged to the set [d, z, n, l, j, ?, h]. Each nonsense word was produced in the constant sentence frame [minu ___ rɑ did] “Minu saw the ___. Using the pitch-synchronous LPC synthesis tools in Multi-Speech Analysis-Synthesis Laboratory, a representative token of each nonsense word was excised from its sentence context, and the constriction interval of the medial consonant was varied by deleting or copying whole glottal pulses at the zero crossings.
Successive members of the same continuum differed by as close to 10 ms as was feasible manipulating whole glottal pulses. Eight variants of each test word were created by these means, through a range that included the mean constriction duration values in the conversational speaking rate for both geminate and singleton versions of the test consonant. The ranges for each class were as in Table 7.

Table 7. Stimulus constriction duration ranges

<table>
<thead>
<tr>
<th></th>
<th>[h]</th>
<th>[ʔ]</th>
<th>[j]</th>
<th>[l]</th>
<th>[n]</th>
<th>[z]</th>
<th>[d]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximum</td>
<td>156.0</td>
<td>148.0</td>
<td>101.5</td>
<td>140.1</td>
<td>149.9</td>
<td>151.6</td>
<td>119.6</td>
</tr>
<tr>
<td>constriction</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>duration (ms)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Minimum</td>
<td>79.7</td>
<td>71.0</td>
<td>27.8</td>
<td>71.4</td>
<td>80.8</td>
<td>82.1</td>
<td>50.5</td>
</tr>
<tr>
<td>constriction</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>duration (ms)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The preceding vowel was set in all tokens to a duration as close to 100 ms as was possible through digital editing, and the following vowel as close as possible to 170 ms. The vowel-consonant transition was kept as it was in the precursor recording in all tokens, with all manipulation limited to the steady-state constriction interval.

3.1.3 Procedure

The stimuli were blocked by consonant and presented in randomized order within block on a computer using SuperLab (Cedrus Corporation). Subjects were asked to indicate by pressing a button on a response pad whether they heard the medial consonant as the simple consonant, indicated in Persian orthography without the length (tashdid) symbol, or as the long consonant,
marked by the same consonant symbol with the *tashdid* marker. Both responses and response times were collected.

7 test words with different medial consonants were included in the study, and each test word was represented by 8 variants differing in medial consonant constriction duration. Each item was heard by each subject 12 times. The total number of stimuli per subjects was thus $7*8*12 = 672$.

### 3.2 Results

Fig. 3 represents the percentage of responses for each stimulus that were geminate, as a function of the stimulus step and the consonant type of the medial consonant.
The lower stimulus steps (to the left) are those with the shortest constriction duration values. In general, the percentage of geminate responses rises from close to 0% for short constriction durations to close to 100% for long constriction durations, confirming that constriction duration is an important acoustic cue for consonant length identification (Garnes 1976, Pind 1986, Hankamer, Lahiri & Koreman 1989, Esposito & Di Benedetto 1999). The darkest plotting characters (the circles) represent the responses to obstruents, and it can be seen that this is the most categorical: it starts off closer to 0 and ends up closer to 100% than the other lines, and it
has a steeper rise in between, with a shorter period of indeterminacy. The other consonant classes have a flatter curve, indicating more equivocal identification as a function of variation in constriction duration.

Each subject's responses for each stimulus continuum were modelled as a cumulative Gaussian distribution (McKee, Klein & Tiller 1985; Wichmann & Hill 2001a,b), using the nls (nonlinear least squares) package in R. This function took the proportion of geminate responses for each subject for each stimulus, and fit it to a psychometric function. One important value of these models is the threshold, which represents the cross-over point in the constriction duration continuum from singleton judgements to geminate judgements. Another important model measure is breadth (also known as the width or spread), which is the standard deviation of the function, and represents the steepness of the cross-over portion of the identification function. The breadth gives the slope of the psychometric function, being greater for flatter slopes. A low value for breadth thus reflects higher sensitivity of listener identification judgements to variation in the manipulated parameter, and a higher value reflects lower sensitivity. Higher breadth values are therefore expected to positively correlate with response times, since response times are also expected to be greater when the task is more difficult.

The mean values for threshold, breadth and response time for each consonant class, pooled across participants, are given in Table 8.
Table 8. Mean breadth, threshold and response times by consonant class (ms)

<table>
<thead>
<tr>
<th></th>
<th>[h]</th>
<th>[ʔ]</th>
<th>[j]</th>
<th>[l]</th>
<th>[n]</th>
<th>[z]</th>
<th>[d]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Threshold</td>
<td>113.4</td>
<td>83.3</td>
<td>61.9</td>
<td>107.6</td>
<td>110.6</td>
<td>112.5</td>
<td>83.3</td>
</tr>
<tr>
<td>Breadth</td>
<td>16.7</td>
<td>20.5</td>
<td>9.4</td>
<td>11.8</td>
<td>9.7</td>
<td>6.4</td>
<td>6.8</td>
</tr>
<tr>
<td>Response time</td>
<td>264.2</td>
<td>305.6</td>
<td>276.9</td>
<td>267.8</td>
<td>266.1</td>
<td>254.8</td>
<td>251.5</td>
</tr>
</tbody>
</table>

The threshold value of the model reflected the inherent duration of the different consonant classes. The highest threshold values were for [h] and [z], which were also the consonant types with the greatest mean constriction duration (103 ms and 95 ms, respectively). The lowest threshold value was for [j], which also had the lowest mean constriction duration of the segments (49 ms).

As expected, the highest breadth values were for the laryngeals, representing a shallow slope to the identification function and a broad zone of equivocation in identification. The lowest values were for the obstruents. Surprisingly, the glide [j] had a lower breadth value than [l] or [n], suggesting that the slope for the glide was in the same range as that for the other sonorants.

A mixed-model ANOVA analysis was run with breadth as a dependent variable, the consonant classes (Obstruent, Glide, Laryngeal) as fixed factors, and subject as random factor. There was a significant main effect of Laryngeal (t = 4.4, p < .0001) and Obstruent (t = -2.3, p = .02) but no effect of Glide (p = .54). The slope of the length identification curve was significantly steeper in obstruents than in sonorants, and significantly less steep in laryngeals than in oral consonants.
The other measure of the ease with which subjects identify the length category of a consonant was the response time, which was expected to be longer where the decision is more difficult. Mean response time was longest for the laryngeal [ʔ], and shortest for the obstruents [z] and [d]. In an ANOVA analysis with the same structure as that for breadth, there were significant main effects of Laryngeal (t = 2.6, p = .009), and Obstruent (t = -2.0, p = .049), but not of Glide (p = .21).

Thus in both breadth and response time, laryngeals were found to be more difficult to assign to length categories than oral consonants. Glides, however, were not found to have a greater breadth or higher response time than other consonants. One reason for this might be the important role of the formant transition duration for glides, which had by a considerable margin the longest mean transition durations of any of the consonant classes (Table 4). Formant transitions were held constant in the stimuli in this experiment, so if this is an important cue for the length contrast for glides, this might have led to less variation in response to glide stimuli than would be the case if the transitions were varying in a more natural manner. The next experiment was designed to specifically test the effect of such variation in formant duration on the perception of length in glides.

4 Experiment #3: Effect of formant transition duration on length class identification

Myers & Hansen (2005) presented Finnish-speaking listeners with stimuli including a glide-vowel sequence in which the vowel formant steady state and the glide-vowel formant transition varied in duration. Subjects had the task of identifying the vowel as short or long. Stimuli with longer formant steady states were more likely to be identified as long, and also
vowels with longer transitions to the glide. They concluded that listeners counted the formant transition as a weighted component of the vowel duration in determining whether the vowel was long or short.

In Experiment 3, stimuli varied in both steady state duration and transition duration, as in the Myers and Hansen (2005) experiment, but in this experiment the steady state that was varied was the constriction interval of an intervocalic glide, and listeners had the task of identifying that glide as long or short. If formant transitions between a vowel and consonant contribute to the perception of length for both vowel and consonant, then we would expect that longer transitions would lead to more identifications of the consonant as long.

Moreover, if, in accordance to our initial hypotheses, longer transitions blur the boundary between neighboring segments, we should expect that longer transitions would be associated with greater uncertainty in length identification, as reflected in higher mean breadth values and response times.

4.1 Methods

4.1.1 Participants

16 adult native speakers of Persian participated in the study: 8 female and 8 male. All were literate in standard literary Persian as used in Tehran, and all were living in the U.S. at the time of the experiment. 8 of the subjects had previously participated in Experiment 2. One of the male subjects failed to complete the experiment due to an equipment failure. Data from two of the male subjects were eliminated from the analysis: one because it turned out he had hearing
deficits, and one because it became clear in the post-experiment debriefing that he had not understood the instructions. These exclusions left data from 13 subjects for analysis.

4.1.2 Stimuli

The stimuli were all based on a production of the nonsense word [qæjːɑb] drawn from the materials for Experiment 2. In all stimuli, the [æ] vowel had a formant steady state lasting 40 ms, and the [ɑ] vowel had one lasting 209 ms. The steady-state portion of the medial glide was manipulated in duration using digital editing, as in Experiment 2, producing a range of 10 equally-spaced constriction duration values from 20 to 110 ms.

The transitions from the preceding [æ] to [j] and from the glide to the following [ɑ] were manipulated using the ASL synthesis package in Kay Elemetrics Multi-Speech. The duration of the interval between vowel and consonant in the precursor was manipulated, and a straight-line interpolation supplied the formant values for the transition. There were three categories of stimuli with regard to duration of the transition between the vowel steady states and the glide steady state, differing by about 20 ms from one step of the series to the next, as summarized in Table 9:
Table 9. Formant transition duration in stimuli

<table>
<thead>
<tr>
<th>Stimulus series</th>
<th>Duration of the [æ] to [j] transition (ms)</th>
<th>Duration of the [j] to [a] transition (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Short</td>
<td>57</td>
<td>88</td>
</tr>
<tr>
<td>Medium</td>
<td>78</td>
<td>108</td>
</tr>
<tr>
<td>Long</td>
<td>98</td>
<td>128</td>
</tr>
</tbody>
</table>

4.1.3 Procedure

There were 30 stimuli in total (10 duration steps for the glide steady-state in 3 transition duration groups), and each stimulus was presented 12 times in a randomized order, for a total of 360 events per subject. The items were presented over headphones on a computer using Superlab (Cedrus Corporation). Subjects were asked to judge whether the medial consonant was long or short, i.e. whether it should be written with the *tashdid* length marker or not. They indicated their response by pressing one of the two buttons on a response pad. Both their response and the response time were collected for each. Response times were measured from the end of the formant transition from the test consonant to the following vowel, since this was the point at which the subject had the relevant information was to the duration of the consonant.

4.1.4 Hypotheses

If listeners take both constriction duration and formant transition duration into account in determining the length class of a consonant, the perceptual boundary between singleton and
geminate would be expected to be at a shorter constriction duration when the formant transition was longer (i.e. model threshold would be lower).

If longer transitions lead to greater indeterminacy in length class identification, then the model breadth and response times would be expected to be greater when the formant transition was longer. All hypotheses were tested by means of a mixed model logistic regression analyses with transition duration class as a fixed independent variable, and subject as a random factor.

4.2 Results

In Fig. 4, the percentage of pooled identification responses that were geminate are presented as a function of the constriction duration and the transition duration. It can be seen that longer constriction durations are identified more often as geminates. The three curves for the different transition durations differ in the crossover from singleton to geminate, which is at longer constriction durations for the shorter transition durations. It can also be seen that the crossover is steeper for the shorter transition durations.

![Figure 4. Percent identification as geminate as a function of constriction and transition duration](image)
The proportion of geminate responses was also greater for longer formant transition classes: .42 for Short, .53 for Medium, and .67 for Long.

As in Experiment 2, the results were fit to a cumulative Gaussian distribution. In this model, the crossover from singleton to geminate is reflected in the threshold. The mean threshold was higher for shorter transitions: Short (72.3 ms), Medium (61.2 ms), Long (46.1 ms). The Long transition class had a significantly lower threshold value compared to the other classes ($t = -9.1$, $p < .001$), and the Short class had a significantly higher threshold than the other classes ($t = 6.8$, $p < .001$). Thus glides with longer transition durations were significantly more likely to be identified as long than comparable glides with shorter transition durations.

The steepness of the crossover is reflected in the cumulative Gaussian model in the breadth parameter. Pooling across subjects, the mean breadth parameter was greater for longer transition durations: Short (9.3 ms), Medium (10.4 ms), Long (13.8 ms). The difference between the Long class and the others was significant ($t = 3.0$, $p = .005$), but that between the Short class and the others was not ($p = .38$). Thus the slope of the identification function was significantly less steep for stimuli with the longest formant transition compared to those with shorter formant transitions.

Mean response time was longer for longer formant transition classes: Short (564 ms), Medium (578 ms), Long (586 ms). The response time for the Short transition condition was significantly shorter than the other conditions ($t = -2.0$, $p = .04$), but the difference between the Long condition and the others was not significant ($p = .24$).

The results supported the hypotheses. Stimuli with longer transitions between the glide and neighboring vowels were more likely to be identified as long than stimuli with shorter
formant transitions. They were also more difficult for listeners to classify, as reflected in higher breadth values and response times.

5 Conclusion

The production study has provided evidence about the acoustic differences between long and short consonants in Persian. Geminates had a longer constriction duration than singletons across speaking rates and consonant classes, though the difference was greater in noncontinuants (stops and [l]). Other acoustic differences between geminates and singletons depended on the consonant class. The vowel was significantly longer before geminates than before singletons only in [l], [n], [z], and [d]. The formant transition was significantly longer in geminates than in singletons just in the supralaryngeal continuants. The drop in intensity from preceding vowel was significantly greater in geminates than in singletons just in [h], [ʔ], [j], and [d].

The production study also provided measures of how similar the different consonant classes of Persian consonants were to the neighboring vowels. In terms of the intensity profile in time, the glides and nonglide sonorants showed a significantly smaller drop in intensity relative to the preceding vowel than the obstruents or laryngeals. But in terms of spectral continuity, the longest formant transitions were those from the vowel to a following glide, and the smallest consonant effect on formant trajectories was in the laryngeals. Obstruents were the least vowel-like consonants in both dimensions.

Experiments 2 and 3 demonstrated that constriction duration is a perceptual cue for consonant length, with listeners tending to identify consonants with longer constriction durations
as geminates. But while the acoustic differences between geminate and singleton are significant in all consonant classes, those classes differ in how easily listeners classify them into long and short categories. In Experiment 2, obstruents were found to be significantly easier to classify into long and short compared to the other consonant classes, and laryngeals were found to be significantly more difficult to classify.

These results supported the general claim that more vowel-like consonant types would be harder to classify into length categories, but against that there was also the fact that glides were not found to be more difficult in this regard than other sonorant consonants. It was suggested that this was due to the role of formant transition variation in the perception of consonant length, since glides had the longest formant transitions of any consonant class in the production study, and the greatest difference between geminate and singleton in formant transition duration.

Experiment 3 addressed the claim that formant transition duration contributes to the perception of consonant length in glides. It was found that intervocalic glides with longer formant transitions were more likely than glides with shorter formant transitions to be identified as long. This supported the hypothesis that listeners consider formant transition duration in making their decision as to whether a glide is geminate or singleton. It was also found that glides with longer formant transitions had a less steep identification curve and longer response times than glides with shorter formant transitions, supporting the claim that longer transitions make length identification more difficult.

These results have provided evidence for the general claim that more vowel-like consonant types are more difficult to classify into length classes than less vowel-like ones, because the more vowel-like consonant types have more gradual transitions to neighboring vowels that blur the acoustic boundary between consonant and vowel (Kawahara 2007, 2012,
Kawahara, Pangilinan, and Garvey submitted). In the case of glides, the relevant transition is a formant transition, while in the case of laryngeals the relevant transition is one in voice quality. These challenges could lead to a tendency for learners to fail to successfully learn length contrasts in glides and laryngeals, which in turn could account for the cross-linguistic avoidance of long glides and laryngeals noted by such authors as Podesva (2000), Kawahara (2007), and Maddieson (2008).

The ambiguity of these transitions also affects the perception of the length of the vowel. Kavitskaya (2002) suggests that the same kind of segmentation difficulty in vowel + glide and vowel + laryngeal sequences leads to lengthening of the vowel in classic compensatory lengthening. Myers and Hansen (2005) argue that such ambiguity in sequences of a vocoid and a vowel leads to a tendency to identify vowels in such a context as long.

Ohala (1983) argued that similar considerations account for the fact that geminate obstruents are often required in the world’s languages to be voiceless, neutralizing voicing contrasts found in singletons. Such a pattern of voicing neutralization occurs, for example, in Nobiin Nubian (Bell 1971), Tamazight Berber (Penchoen 1973), Endegen (Rose 2006), Japanese (Kawahara 2006), and Moro (Jenks & Rose 2011). Further examples are discussed in Hayes & Steriade (2004). Ohala relates this common pattern of neutralization to the difficulty of maintaining a translaryngeal pressure drop in a long obstruent. Vocal fold vibration ceases when translaryngeal airflow ceases, which increases the difficulty of distinguishing voiced from voiceless geminates. A merger of voiced with voiceless in this context would be a hypocorrective sound change in the typology of Ohala (1993).

The approach can be extended to another common restriction on geminates: that they be word-medial. More specifically, a number of languages have the restriction that geminates occur
only in the context between two vowels, e.g. Norwegian (Kristoffersen 1999: 632), Lango
(Noonan 1992: 11), Malayalam (Syamala Kumari 1972: 48), Tümpisa Shoshone (Dayley 1989:
xxix), and Mokilese (Harrison 1976: 24). The medial intervocalic context is the one best suited
to locating both the onset and the offset of the consonant, and thus it is the easiest position in
which to determine a consonant's duration. The difficulty of hearing and measuring length
contrasts in utterance-initial position, for example, has been emphasized in both transcription-
based and experimental studies, in Luganda (Tucker 1962), Arabic (Obrecht 1965), Pattani
Malay (Abramson 1986, 1999), Tashliyt Berber (Ridouane 2007), and Swiss German
(Kraehenmann & Lahiri 2008). These perceptual difficulties reduce the likelihood of a length
contrast surviving in such contexts, just as the perceptual difficulties of length identification for
vowel-like consonants, explored in this paper, reduces the likelihood of survival for that kind of
contrast.

References

Journal of the International Phonetic Association 16, 8–16.

Abramson, Arthur S. 1999. Fundamental frequency as a cue to word-initial consonant length:
Pattani Malay. In Proceedings of the XIVth International Congress of Phonetic Sciences,
Berkeley, University of California (ICPhS 14), vol. 1, 591–594.


Aoyama, Katsura & Lawrence Reid. 2006. Cross-linguistic tendencies and durational contrasts in
geminate consonants: An examination of Guinang Bontok geminates. Journal of the


