Production of Persian Geminate Stops: Effects of varying speaking rate

Benjamin B. Hansen
University of Texas at Austin

1. Introduction

Tehrani Persian has a phonological contrast between geminate and singleton consonants. For example, [made] (female) contrasts with [mad:e] (material). An interesting observation has been made by Iranian researchers regarding geminates: for certain words speakers may pronounce geminate consonants as a singletons. Mahootian (1997) states that for native (non-Arabic) words "the geminates in these words are often, if not usually, reduced." Deyhime (2000) transcribed the pronunciations of 16 speakers of Tehrani Persian and whereas geminated stop consonants in words of Arabic origin are pronounced as geminates by all informants, the geminate stops in native words are pronounced by some of the informants as singletons. For example, the word for "childhood" is pronounced by some informants as [bačegi], and by others as [bač:egi]. Ohala (1981) describes a process by which "mini sound changes" arise from misclassification of segments when a listener fails to decode the phonetic context and thereby opens the way for phonologizing a context-dependent articulation. Lindblom et al. (1995) suggest that the acceptability of an alternate pronunciation is measured in part by its "social value" in addition to its "articulatory complexity" and its "perceptual distinctiveness." The case of the universally-enforced gemination of stops in Arabic loanwords in Persian could be seen as an instance of social value upholding a canonical pronunciation in the face of a sometimes-ambiguous distinction in real-life speech. This study seeks to identify whether the geminate and singleton categories do in fact overlap at normal and two faster speaking rates.

Several studies have sought to determine the acoustic cues for gemination. Lahiri & Hankamer (1988) and Hankamer et al. (1989) studied gemination in Turkish and Bengali, where they detected the presence of secondary cues from perception tests, but could not identify which parameter other than duration actually influenced the perception of gemination. They concluded that closure duration is the single overriding cue for gemination in both Turkish and Bengali. For Italian stops, Pickett et al. (1999) found that a short preceding vowel duration was a secondary cue for gemination, but the stop closure duration cue is strong enough to override it: only when the closure duration was ambiguous between the geminate and singleton categories did the length of the preceding vowel influenced judgments of the geminate/singleton distinction. Pind (1995) found that for Icelandic, the ratio of the consonant to the preceding vowel was the primary cue for gemination and this ratio remains invariant at different rates. He reports the discriminating ratio to be close to 1. Han (1994) reports that for Japanese, a lengthened preceding vowel may be a secondary cue for gemination, but the observation is not supported by a quantitative analysis.

The idea that the listener may have to factor out rate-dependent changes has also been the subject of a number of studies. Miller (1981) found that listeners adjust to the speaking rate when using the F1 formant transition duration as a cue to distinguish between /b/ and /w/ in English. She concluded that rate-dependent perceptual effects exist. Pickett et al. (1999) found that perception of the geminate/singleton distinction in Italian is conditioned by the speaking rate, but that a higher order relational measure (the closure duration/preceding vowel duration ratio) remains constant as a discriminator between the geminate and singleton categories. This latter result coincides with Pind's observation cited above. Kessinger & Blumstein (1997) on the other hand concluded that an adjustment of the category boundary is not necessary to perceive the distinctions between long lag and short lag voice onset time (VOT) in Thai and English or between prevoiced and short lag VOT in French and Thai; a single discriminating value for each distinction may be applied across speaking rates. Finally, Magen & Blumstein (1993) discovered that in Korean, no asymmetry exists in the response of long vs. short vowels to changes in speaking rate: both categories are equally affected by rate.

A pilot study that I performed on Persian geminates showed that in comparing geminate to singleton durations in isolated and connected speech, different threshold values were required to discriminate between the categories for each style, since the isolated singleton durations overlapped the connected geminate durations. When the data were presented in a scatter plot of the stop closure durations vs. the preceding vowel durations, the two groups of data were shown to be separate. The implication is that a listener may be able to compensate for the apparent blurring of the geminate/singleton distinction, as represented by the closure duration, by factoring out the difference in speaking style, picking up on cues such as the duration of the preceding vowel. In the current study involving different speaking rates, I measured both the consonant duration and the preceding vowel duration, as well as the total utterance length. I am interested in discovering whether at faster speaking rates, geminate durations are so reduced as to fall within the range of singleton durations, and whether such rate-dependent cues as the duration of the preceding vowel and the overall speaking rate may be used to discriminate between the two categories.

2. Methods

Three literate adult speakers of standard Tehrani Persian served as informants for the study. The two female and one male speakers reside in the United States and use Persian regularly in social and family contexts. They have lived outside of Iran for between 6 and 25 years.

The informants read carrier sentences containing 12 two-syllable study words. The words were a mixture of commonplace and more literary words. The study words contained singleton or geminate alveolar stops in an intervocalic environment with word stress on the following syllable. Each of the alveolar stops [t,d] occurs three times in singleton form and three times in geminate form. Half of the singletons and half of the geminates were preceded by the low vowel [æ]. The remaining stops were preceded by a mid vowel [e,o]. The high vowels [i,u] are rare preceding geminates in Persian, so they are not present in the preceding vowel for these data.

The study words are listed below:

qætar	*train*
ketab	*book*
šotor	*camel*
mætte	*drill*
fættan	*temptor*
botte	*bush*
fæda	*devotion*
sædæf	*shell*
medad	*pencil*
mæddah	*eulogist*
šeddæt	*hardship*
moddæt	*time period*

The carrier sentence for each word was:

Færhæng goft ke ______ næbud.
Farhang said that ______ not-was.
Farhang said that (the) _________ wasn't there.

The informants were first asked to read the sentences at a self-selected comfortable rate, so as to be clearly understood. I referred to this speaking rate as "normal". Next, they were asked to read the sentences faster, but to make sure that the words were still understandable. I called this rate "fast." Finally, the informants were asked to read the sentences very much faster, without regard to whether the words were clearly understandable. I called this speaking rate"fastest." Three repetitions of the sentences were recorded for each speaking rate, resulting in 108 tokens per speaker, or 324 total for the experiment.

The recordings were made in the acoustically isolated sound recording room in the Linguistics Laboratory at the University of Texas at Austin with a Sure BG-3.1 microphone. The tokens were recorded on a Fostex D-5 digital tape recorder and then digitized using a sampling rate of 11025/sec. I used Macquirer acoustic analysis software to produce a waveform display and spectrogram for each token.

For each token I measured three duration parameters associated with the medial stops: preceding vowel duration, closure duration and utterance duration.

Preceding vowel duration: Using the waveform display I marked the onset of the vowel at the zero amplitude crossing preceding the first period of the waveform exhibiting the characteristic periodicity of the vowel as signaled by the presence of a formant structure appropriate to the vowel. The vowel offset was marked at the zero amplitude crossing of the waveform following closure of the stop. The stop closure is marked by an abrupt decrease in amplitude of the waveform. The spectrogram display was used as a reference to identify the segments.

Closure duration: The closure duration was measured from the zero amplitude crossing at the offset of the preceding vowel, to the onset of the burst at the release of the stop. The burst is identified by the abrupt occurrence of high frequency energy following closure.

Utterance duration: The total utterance duration was measured from the onset of the high-frequency frication characterizing the first segment [f] of the carrier sentence, to the burst accompanying the release of the final stop [d]. In some of the faster tokens there was not a perceptible burst associated with the final stop. In these cases I measured the utterance duration up to the point at which voicing ceased.

I used StatView statistical software for statistical analysis and to generate box plots, Excel for scatter plots and regression analysis, and ViSta for 3-D visualization plots.

3. Results

Table 1 presents summary of the average values obtained for the parameters Consonant Closure (C), Preceding Vowel (V), and Consonant-Vowel Ratio (C/V). Results are presented for each informant and for the pooled totals. The average ratio of the geminate to singleton durations is also shown for each speaking rate. While the same overall pattern of decreasing values at faster rates is apparent for all three parameters, individual differences can be observed. Speaker F-2 had much shorter singleton closures than the other two speakers, resulting in higher Geminate/Singleton ratios and lower C/V ratios for singletons at each speaking rate. Speaker M1 had consistently shorter vowels than F-1, and F-1 had shorter vowels than F-2. For all speakers, the ratio of geminate to singleton closures consistently decreased with increased speaking rate.

F-1

	C (ms)			V (ms)		C/V	
	Gem.	Sgl.	Gem./Sgl.	Gem.	Sgl.	Gem.	Sgl.
Normal	156.5	73.4	2.13	121.2	102.2	1.32	0.74
Fast	110.1	56.8	1.94	99.6	88.1	1.13	0.67
Fastest	87.6	45.5	1.92	80.8	72.0	1.11	0.69

F-2

	C (ms)			V (ms)		C/V	
	Gem.	Sgl.	Gem./Sgl.	Gem.	Sgl.	Gem.	Sgl.
Normal	148.3	46.5	3.19	132.6	109.8	1.12	0.44
Fast	119.8	42.5	2.82	116.1	98.3	1.04	0.45
Fastest	89.7	35.4	2.53	92.9	81.1	1.00	0.46

M-1

	C (ms)			V (ms)		C/V	
	Gem.	Sgl.	Gem./Sgl.	Gem.	Sgl.	Gem.	Sgl.
Normal	155.8	71.1	2.19	112.1	98.1	1.29	0.83
Fast	120.2	55.8	2.15	95.9	83.1	1.27	0.75
Fastest	84.9	51.2	1.65	82.7	74.2	1.06	0.79

POOLED

	C (ms)			V (ms)		C/V	
	Gem.	Sgl.	Gem./Sgl.	Gem.	Sgl.	Gem.	Sgl.
Normal	153.5	63.7	2.41	122.0	103.4	1.24	0.67
Fast	116.7	51.7	2.25	103.9	89.8	1.15	0.62
Fastest	87.4	44.1	1.98	85.5	75.7	1.05	0.65

Table 1. Summary of mean duration measurements and ratios. N=18 for individual informant data. N=54 for pooled data.

Figure 1 is a box plot representing the distribution of consonant closure durations, pooled for all three speakers. The decrease in the geminate durations is much more dramatic than the decrease in singleton durations, although the latter also decrease consistently with rate. Figure 2, representing pooled preceding vowel durations, shows that both before geminates and singletons, the vowel decreases with increased rate. Vowels preceding singletons are consistently shorter than vowels preceding geminates at the corresponding rate, though much of the range of values overlaps. The pooled C/V distributions are shown in Figure 3. There is considerable overlap among the categories across rates, but there is a consistent decrease in the parameter with increased speaking rate, that seems to be more appreciable for geminates than for singletons.

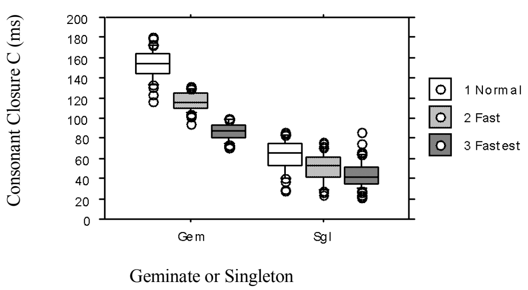

Figure 1. Box plot of consonant durations (C) for three self-selected speaking rates. Pooled data from three speakers.

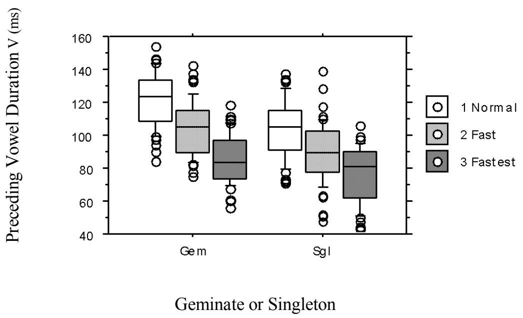

Figure 2. Box plot of preceding vowel durations (V) for three self-selected speaking rates. Pooled data from three speakers.

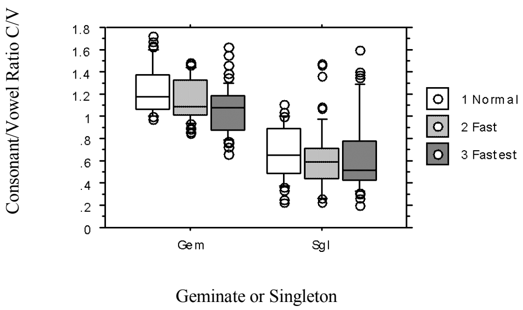

Figure 3. Box plot of consonant to vowel ratio (C/V) for three self-selected speaking rates. Pooled data from three speakers.

I performed ANOVA analyses of the three parameters (C, V, C/V) for the 324 pooled tokens to determine whether there is a significant interaction between speaking rate and the parameter gemination (i.e. the parameter that indicates whether the consonant is a geminate or not). The results show that there is a significant interaction between speaking rate and gemination for the consonant closure C ($F=89.379$, $p<0.0001$). This result can be interpreted to mean that increased speaking rate has a greater influence on the duration of geminates than it does on the duration of singletons, and is consistent with the observation that in Figure 1 the geminate duration decreases more dramatically.

The analysis for the preceding vowel V did not indicate that there is significant interaction between speaking rate and gemination ($F=1.876$, $p=0.1548$). This implies that vowels shorten with increased rate preceding geminates much as they do preceding singletons. The rate*gemination interaction for the consonant-vowel ratio C/V is marginally non-significant at the assumed significance level of $p=0.05$ ($F=2.987$, $p=0.0519$). As can be seen in Figure 3, the C/V ratio does seem to decrease more consistently for geminates, but due to the considerable overlap between speaking rates, the effect is not as clear-cut as is seen with the closure durations alone. In general, however, it can be stated that both C and C/V are more susceptible to being reduced at faster rates when they are geminates. This result is consistent with the observation made by Pickett et al. for Italian: "geminate distributions appear to be more affected by increased speaking rate than the singleton categories."

A scatter plot of the consonant closure against the preceding vowel duration serves to clarify the interactions between the parameters C and V. Figure 4 shows the data obtained from informant F-1. In the figure the solid symbols represent the geminate values. The cluster of values for each speaking rate is circled with a solid line for geminates. Singleton values are represented by hollow symbols and the data from each rate are circled with dashed lines. Although there is overlap between the geminate and singleton values, there is no overlap between the categories *within a particular speaking rate*. Thus, even though a singleton token produced at the normal rate may fall within the range of the values observed for the fastest geminates, a singleton token produced at the fastest rate does not overlap the fastest geminates. Similar plots from the other informants show the same general relationships.

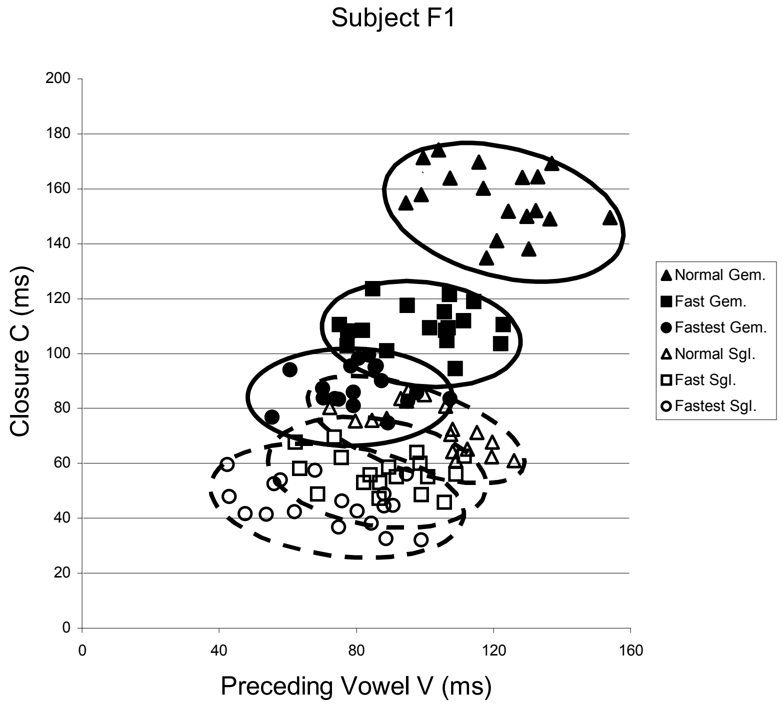

Figure 4. Scatter plot of consonant closure durations (C) vs. preceding vowel durations (V) for geminates and singletons produced by informant F-1 at three self-selected speaking rates. Solid symbols are geminates, hollow symbols are singletons. Oval represent clusters of data produced at a self-selected selected speaking rate. N=18 within each cluster.

It is notable that while the *mean* C/V ratios are greater than 1 for geminates and less than 1 for singletons for all informants, Figure 4 shows that no single C/V ratio is can be found to cleanly discriminate between the clusters of geminates and singletons. This observation is quite unlike Pind's and Pickett et al.'s conclusions for Icelandic and Italian. In those languages C and V durations were shown to be complementary, with geminates being preceded by shortened vowels. Their plots of C vs. V showed the geminates to be completely separate from singletons and a single ratio could be selected to almost completely define the boundary between them. In the Persian data, it appears that the distinction can be better defined by a threshold consonant duration, but that a different threshold obtains for each speaking rate.

As noted above, in Persian, the preceding vowel is *lengthened* before geminates, as was reported to be the case for Japanese by Han. This positive correlation between C and V makes it difficult to see that the categories are in fact distinct if speaking rate is taken into account. However, the separation can be made apparent by including speaking rate as a third dimension to the plot. I expressed speaking rate as the average syllable length in milliseconds, equal to the utterance duration divided by 8 (each carrier sentence containing a test word was 8 syllables long). Figure 5 shows a 3-D scatter plot of the pooled data from all three informants. The three axes are consonant closure C, preceding vowel V and average syllable length, all in milliseconds. Geminates are shown as solid squares and singletons as hollow squares. With one or two exceptions, there is no overlap between the categories in this plot.

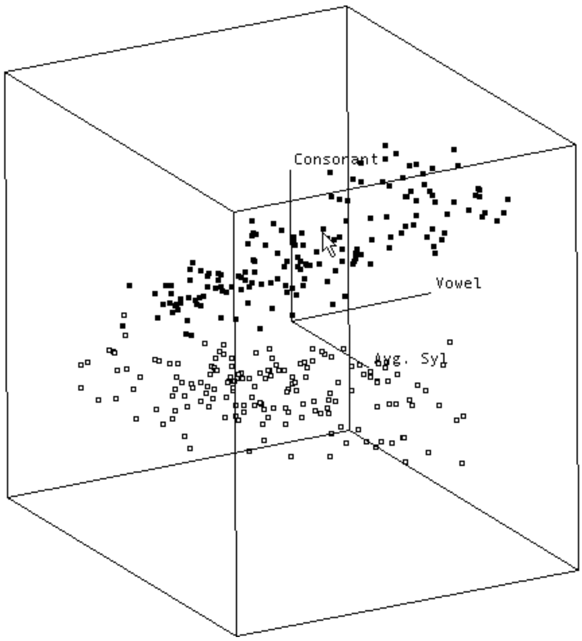

Figure 5. 3-D scatter plot of consonant closure duration (ms) vs. preceding vowel duration (ms) vs. average syllable duration (ms). Geminates are solid squares (N=162), singletons are hollow squares (N=162). Pooled data from three informants.

What is unclear at this point is what parameter best serves to distinguish the geminate and singleton categories. A plot of consonant closure vs. average syllable duration from informant F-1 (Figure 6) shows the categories to be distinct at each speaking rate, but no single

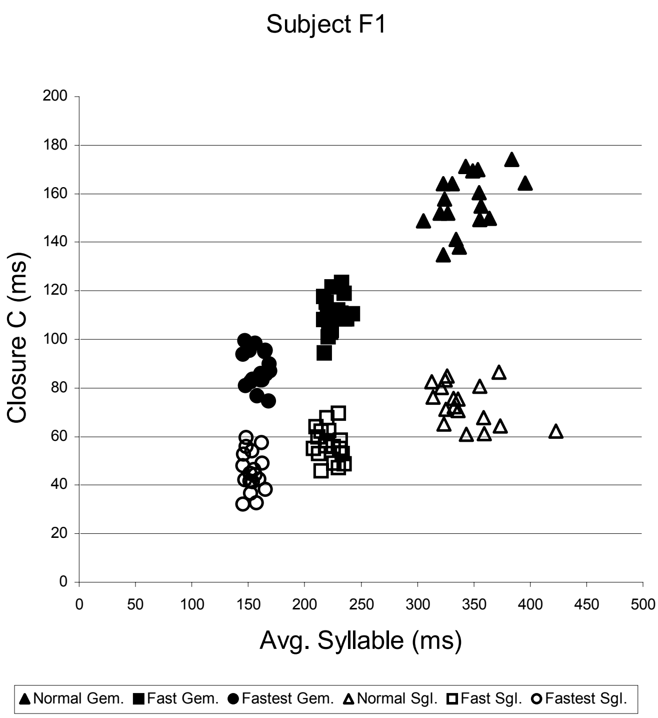

Figure 6. Scatter plot of consonant closure duration vs. average syllable duration. Solid symbols are geminates, hollow symbols are singletons. Data are from informant F-1.

threshold value or slope can be found to define the category boundary. An additional observation can be made from this plot: both the geminate and singleton trends appear to intercept the ordinate at about 20 ms.

On the hypothesis that there may be a minimum closure duration for stops as was posited by Klatt (1973), I observed that after subtracting a "minimum" duration value from both geminates and singletons, the durations appeared to increase by power functions of the rate (as expressed by the average syllable length). I found that if the "minimum" is set to 17.9 ms for informant F-1, for example, the exponent of the two power functions representing geminates and singletons equals 0.89 for both functions. When two power functions having the same exponent are plotted on a log-log scale, the functions appear as parallel lines as is shown in Figure 7. The equations of the lines are of the form:

$$y = kx^{\beta}$$

where y is the marginal duration above the minimum duration C_0 ($y = C - C_0$), k is a constant, x is the average syllable duration, and β is the exponent of the power function. The geminate-singleton distinction is then in the parameter k: For informant F-1, k = 0.30 for singletons and 0.75 for geminates. A single value for k, say, k = 0.6, is sufficient to completely distinguish the geminate and singleton categories for this speaker. k may be understood as a fixed proportion of an adjusted average syllable duration. A geminate would be a stop whose duration exceeds a particular proportion of the average syllable, above the minimum duration. The corresponding analysis for informant F-2 gave values of 17.4 for C_0 and 0.92 for β. For informant M-1, the values were 29.2 for C_0 and 1.31 for β.

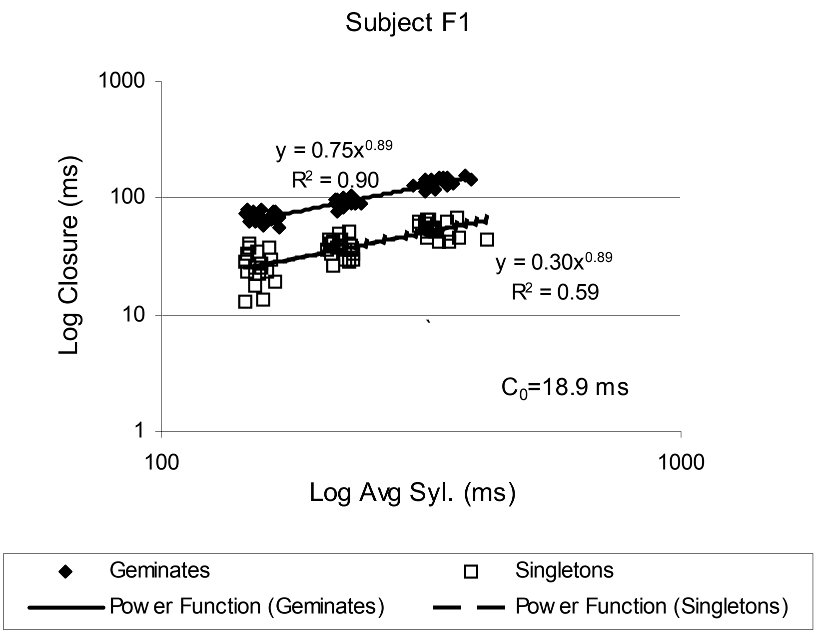

Figure 7. Log-log plot of the same consonant closure vs. average syllable duration data shown in Figure 6, adjusted by subtracting a minimum closure value C_0 = 18.9 ms. Data are from informant F-1.

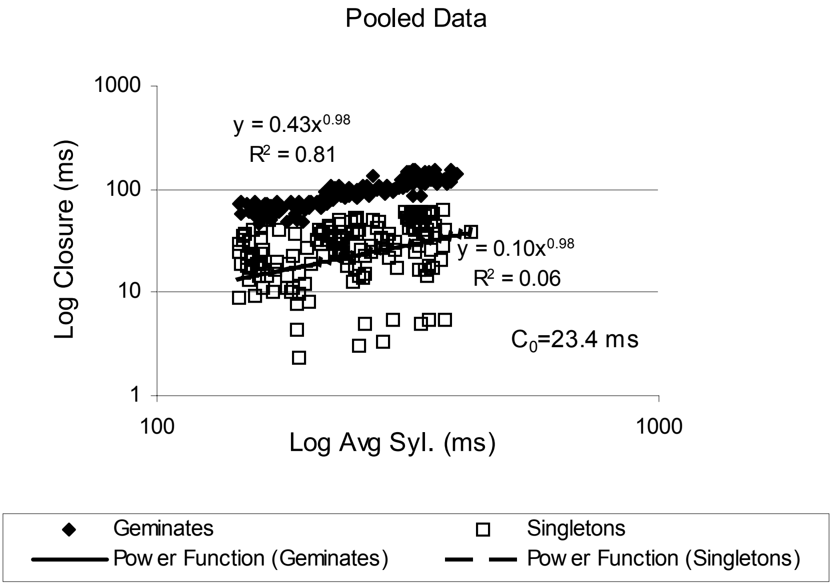

Figure 8. Log-log plot of the pooled consonant closure vs. average syllable duration data. Consonant closure data are adjusted by subtracting a minimum closure value C_0 = 23.4 ms.

In the pooled data, when $C_0 = 23.4$ is assumed, $\beta = 0.98$ for both regression lines as shown in Figure 8. Note that the constant value k is sensitive to the assumed C_0 value. For the pooled data, the threshold k value would be on the order of 0.35.

It is interesting to note the similarity between the form of the power equation used to describe these data and the power equation assumed by Stevens (1986) to describe psychophysical phenomena. Stevens' psychophysical power law is:

$$\psi = k\varphi^{\beta}$$

where ψ is the estimated perceived magnitude, k is a calibrating constant, φ is a scalable stimulus and β is a characteristic exponent for the type of phenomenon being perceived. For example, Stevens found that people's estimates of loudness magnitude obeyed a power law having β equal to 0.67. For visual length, the exponent was 1.0, for white noise duration, 1.1, and for muscle force, 1.7, among many other examples.

If we look at the speaking rate as a stimulus and the consonant closure as an estimate of perceived duration, we could say that consonant duration obeys a psychophysical law: the consonant is held just so long as needed to be perceived by the speaker as being geminate or singleton.

4. Conclusions

Persian geminate stops are clearly distinguished from singleton stops in production when speaking rate is taken into account. Although there is overlap between the categories as observed across all speaking rates, there is almost no overlap within a speaking rate, even for different speakers. The pattern observed in Persian is distinct from that reported for Italian and Icelandic. The C/V ratio is not a clear discriminator between the categories as has been observed for those languages. This is partly due to the fact that in Persian the vowel preceding the geminate tends to be longer. The average syllable duration is a more useful parameter for understanding the distinction between the categories. I found that after subtracting a minimum closure duration, the stop closure varies as a power function of the average syllable duration. However, the observation that a single exponent applies to both geminates and singletons is not a strong one. The exponent is highly sensitive to the selected minimum duration. All that can be said is that for each speaker, there is some minimum duration value for which the exponents are the same. There is no independent way to determine what the minimum duration should be. Indeed, the values of C_0 found for the three speakers varied considerably.

The notion that the closure duration obeys a psychophysical power law is an evocative one. However it is not transparently clear that the speaking rate is a scalable stimulus in the psychophysical sense, or that the produced duration is a magnitude estimation. In any case, it is not counter-intuitive to posit that the speaker holds the closure as long as needed to occupy a set perceived proportion of the average syllable, above and beyond the minimum duration required for the production of a stop.

We can therefore tentatively conclude from the production data that an invariant discriminator exists between geminate and singleton categories, and this discriminator applies across speakers and speaking rates. Further perception studies are underway to determine whether the perceived boundary between geminates and singletons at different speaking rates coincides with the boundary observed in these production data.

On the question of why certain words have optional gemination, these data suggest that the phenomenon cannot be accounted for by perceptual confusion, but rather the explanation must be sought in the realms of sociolinguistics or historical and comparative dialectology.

References

Deyhime, G. (2000) *Farhang-i Avayi-i Farsi* (Persian Pronunciation Dictionary) Tehran: Farhang Moaser Publishers

Han, M.S. (1994) Acoustic manifestations of mora timing in Japanese, *Journal of the Accoustical Society of America*, 96, 1, 73-82

Hankamer, J., Lahiri, A. & Koreman, J. (1989) Perception of consonant length: voiceless stops in Turkish and Bengali, *Journal of Phonetics,* 17, 283-298

Kessinger, R. H. & Blumstein, S. E. (1997) Effects of speaking rate on voice-onset time in Thai, French and English, *Journal of Phonetics,* 25, 143-168

Klatt, D. H. (1973) Interaction between two factors that influence vowel duration, *Journal of the Accoustical Society of America*, 54, 4, 1002-1004

Lahiri, A. & Hankamer, J. (1988) The timing of geminate consonants, *Journal of Phonetics,* 16, 327-338

Lindblom B., Guion, S., Hura, S., Moon, S. & Willerman, R. (1995) Is sound change adaptive? *Rivista di Linguistica*, 7, 1

Miller, J. L., (1981) Some effects of speaking rate on phonetic perception, *Phoneteca,* 38, 159-180

Magen, H. S. & Blumstein, S. E. (1993) Effects of speaking rate on the vowel length distinction in Korean, *Journal of Phonetics,* 21, 387-409

Mahootian, S. (1997) *Persian*. London: Routledge

Ohala, J. J. (1981) The Listener as a Source of Sound Change, In *Papers from the Parasessions, Chicago Linguistic Society*, May, 178-203

Pickett, E. R., Blumstein, S. E. & Burton, M. W. (1999) Effects of Speaking Rate on the Singleton/Geminate Consonant Contrast in Italian, *Phonetica,* 56, 135-157

Pind, J. (1995) Speaking rate, voice-onset time, and quantity: The search for higher-order invariants for two Icelandic speech cues. *Perception & Psychophysics*, 57, 291-304

Stevens, S. S. (1986) *Psychophysics*. New Brunswick: Transaction

Proceedings of the 2003 Texas Linguistics Society Conference: Coarticulation in Speech Production and Perception

edited by Augustine Agwuele, Willis Warren, and Sang-Hoon Park

Cascadilla Proceedings Project Somerville, MA 2004

Copyright information

ISBN 1-57473-402-4 library binding

Ordering information

Orders for the library binding edition are handled by Cascadilla Press.
To place an order, go to www.lingref.com or contact:

Cascadilla Press, P.O. Box 440355, Somerville, MA 02144, USA
phone: 1-617-776-2370, fax: 1-617-776-2271, e-mail: sales@cascadilla.com

Web access and citation information