

## CHAPTER 4: EXPERIMENTS 2 AND 3 (PERCEPTION EXPERIMENTS)

### 4.1. Introduction

The hypothesis that velar palatalization arises from a perceptual reanalysis of velars before front vowels in faster speech makes several testable predictions: (1) velars before front vowels will be acoustically similar to coronals, especially the voiceless velars, (2) that similarity will be heightened in faster speech, and (3) fronted velars will be perceptually confusable with coronals. The first two predictions are supported by the acoustic analysis discussed in Chapter 3. First, the velars before front vowels were found to be more like the palatoalveolar affricates than velars before back vowels in terms of the spectral characteristics of the consonantal release and the F2 transitions. Secondly, the velars and palatoalveolars were found to be more similar before the high front vowels in terms of their spectral characteristics. Thirdly, the voiceless velars were more like palatoalveolars than voiced velars in terms of the consonantal spectral characteristics, the duration of the consonant, and the following F2 transition. Fourth, the similarity between fronted velars and palatoalveolars was greater in faster speech than in citation speech in terms of the spectral characteristics of the consonantal release. The third prediction, that fronted velars will be perceptually confusable with coronals, is tested in this chapter.

This chapter reports the results of two perception experiments, Experiments 2 and 3, in which subjects were asked to identify tokens of [k], [tʃ], [g], and [dʒ].

Tokens of [k] were often heard [tʃ], especially when the [k] was before a high front vowel. The acoustic similarity between [tʃ] and [k] before front vowels discussed in the Chapter 3 is reflected in the perception experiments.

The prediction tested in Experiments 2 and 3 can be broken down into three parts. The first prediction is that by degrading the signal we will produce [k]/[tʃ] confusion with a vowel asymmetry whereby [k]'s before high front vowels are more often heard as [tʃ] than before back vowels. The second prediction is that faster speech productions of [k] will more often be heard as [tʃ] than the citation productions. The third prediction is that there will be more [k]/[tʃ] confusion than [g]/[dʒ] confusions based on the acoustic as well as typological evidence that voiceless velars are more likely to undergo palatalization than voiced velars (see §2.2).

The signal was degraded in two ways in the experiments discussed. Experiment 2 uses gating to shorten the consonant and Experiment 3 uses noise to mask the signal. In general, they support the predictions outlined above. Both show misperceptions of [k] as [tʃ] and have the predicted vowel asymmetry effect. The results from the gating experiment indicate that [ki] sequences are misheard as [tʃ] about half the time. The [kɑ] and [ku] sequences on the other hand, are heard correctly over 90% of the time. The results from the noise masking experiment also have a vowel asymmetry effect. In this case, the [ki] and [ku] sequences are more often heard as [tʃ] than the [kɑ] sequences. In addition, there is a greater confusion rate of [k] for [tʃ] than of [g] for [dʒ]. The results from the noise masking experiment do not seem to support the prediction that velar tokens from faster speech are more often misheard than velar

tokens from citation speech. It is suggested that the gross effects of the noise obviate the more subtle faster/citation distinction.

Two acoustic attributes, VOT and peak spectral frequency of the burst, are also investigated. Velar stops with higher peak spectral frequencies and longer VOTs are more often identified as palatoalveolars. The [dʒ] stimuli are also often heard as voiceless. This is a surprising finding since voicing is usually found to be one of the most robust features in consonant confusion experiments (See e.g., Miller and Nicely 1955). An acoustic attribute was found to have an effect on the identification: the [dʒ] with longer durations are heard as voiceless.

## 4.2. Experiment 2: Gating

In this section, I present the results of a perception experiment in which [k] and [tʃ] consonants have been shortened to 30 ms. and then presented to listeners for identification. The results suggest that the acoustic similarity between [k] and [tʃ] has a perceptual correlate: [k]'s before front vowels are confused with [tʃ].

### 4.2.1. Methodology

The stimuli consisted of tokens of [k] and [tʃ] taken from a context before [i], [u], and [ɑ]. The tokens were the same as those used in the acoustic experiment. The words used were: *quiche*, *cot*, *coop*, *chief*, *chop*, *chew*. The stimuli were collected from the faster speech of two speakers, one female (RG) and one male (NS). There were four repetitions of each word type by each speaker. Thus, there were 8 different tokens for each of the 6 types, making a total of 48 tokens. These forms were then

digitally edited in two ways. In the first, information following the consonant was removed so only the burst and aspiration was left in the case of [k] and only the release and frication was left in the case of [tʃ]. In the second editing, only the first 30 ms. of the consonant was kept and the rest of the token was deleted. All of the consonants were more than 30 ms. long, so no vowel information remained. These tokens were randomized and then played to 19 listeners. The longer condition was presented first, followed by the truncated condition. The listeners were asked to decide whether they heard a [k] or a [tʃ] in a forced choice. There were a total of 1824 responses, 912 responses per condition.

#### 4.2.2. Results

Table 4.1 shows the results of the first condition in which all the aperiodic noise was retained. The numbers given are percentages. The diagonal line running from top left to bottom right shows the percent correct identifications. If all the responses were 100% correct, there would be 100% in all the boxes in this diagonal. Note that the only mistakes made were for the [k] before the high front vowel [i], but that for the most part the identification was at or near 100% correct.

## All Consonantal Information

		SPOKEN					
		[k]			[tʃ]		
H E A R E D		[i]	[ɑ]	[u]	[i]	[ɑ]	[u]
	[k]	98%	100%	100%	*	*	*
	[tʃ]	2%	*	*	100%	100%	100%

Table 4.1

Now consider the results for the second condition (in Table 4.2) in which only the first 30 ms. of the tokens were played. Note that [k] before [ɑ] and [u] and [tʃ] before [i], [ɑ] and [u] were all identified correctly more than 90% of the time. The [k] before [i] on the other hand was only identified correctly 53% of the time. This is right around the level of chance. The tokens beginning with [k] and the high front vowel [i] are highly confusable with the palatoalveolar affricate. The subjects appear to have been guessing between a response of [k] and [tʃ] for these tokens.

## First 30 ms. of Consonant

		SPOKEN					
		[k]			[tʃ]		
H E A R D		[i]	[a]	[u]	[i]	[a]	[u]
		[k]	53%	97%	97%	*	2%
	[tʃ]	47%	3%	3%	100%	98%	94%

Table 4.2

The results from Experiment 1 tend to support the prediction that [k]'s before front vowels will be confusable with [tʃ]. It is interesting to note that the confusion is uni-directional. [k]'s are heard as [tʃ], but [tʃ]'s are not heard as [k]. This directionality parallels the typological observation that velars palatalize to palatoalveolars, but that palatoalveolars do not become velars before front vowels. It is unclear however, whether the [ki] sequences sound like [tʃ], or if the shortened duration simply renders the consonant unintelligible. Since there are only two choices in the paradigm, we do not know if the 47% [tʃ] responses to [ki] are due to random identification of an unknown segment, or if the [ki]'s do indeed sound like [tʃ].

### 4.3. Experiment 3: Noise Masking

The results from the Experiment 2 were quite promising and indicate that [k] before high front vowels could be confusable with [tʃ]. However, a methodology with more than two responses is needed to see if the results were due to random guessing or

genuine [k]/[tʃ] confusion. A larger experiment is also needed to investigate the role of voicing and speech style on the consonant identification.

In this section, I report on a larger experiment (Experiment 3) which investigates the [k]/[tʃ] confusion further. There are four possible responses and speech style and voicing have been factored into the design. In this experiment the signal was degraded by using masking noise instead of shortening the stimulus. This change was prompted by a couple of reasons. First, it was hoped that the noise condition would more closely model real world listening conditions. Noisy conditions are quite common in everyday conversation whereas sounds truncated to the first 30 ms. are more rare. However, the noise level needed to induce a large number of perceptual errors was quite exaggerated compared to real world conditions. So, it is unclear how comparable the noise masking condition is to real world listening conditions. Second, the noise masking condition, unlike the gating condition, allows investigation into durational cues such as VOT.

The perception tests discussed here are unlike those in the literature in that subjects are given the response choices of stop and affricate, and also that the responses are analyzed in terms of vowel context. Studies such as Repp and Lin (1989) and Winitiz *et al.* (1972) have looked at stop confusions in terms of vowel context. They found that a velar stop before a front vowel is often inaccurately perceived as a coronal stop. In these studies affricates were not a possible answer. Ahmed and Agrawl (1969) and Wang and Bilger (1973) have conducted perception experiments which had many manners and places of articulation, including stops and affricates. In these experiments, however, the responses were not analyzed in terms of vowel context. Ahmed and Agrawl looked at Hindi consonants and found that

affricates were the least confused segments. In other words, the affricates were the most perceptually salient consonants. Wang and Bilger also found that [tʃ] was more often correctly identified than [k].

The results from this study support the prediction that [k] and [tʃ] will be quite confusable, especially before high vowels. The results also agree with the typological evidence that [k] to [tʃ] palatalization is more common than [g] to [dʒ] palatalization in that [k] is heard as [tʃ] more often than [g] is heard as [dʒ]. Also, in agreement with previous studies (Ahmed and Agrawl 1969, Wang and Bilger 1973) the palatoalveolars have an overall higher identification rate. Evidence is also presented that faster speech [g]'s are heard as [j], a finding consistent with the observation that [g]'s often palatalize to [j].

#### 4.3.1. Methodology

The stimuli for Experiment 3 were taken from faster and citation productions of words beginning with the consonants [k tʃ g dʒ] followed by the vowels [i ɑ u]. The tokens were the same as those used in Experiment 1 in the acoustic investigation. The words used were as follows: *quiche, cot, coop, chief, chop, chew, geese, got, goose, jeep, jot, juice*. Seven speakers (four female and three male) produced four tokens of each word in faster and citation speech. (See chapter 2 for a detailed account of how the data were collected). Thus there were 24 types spoken four times each by 7 speakers for a total of 672 tokens in the listening task. 20 subjects heard the tokens with masking noise for a total of 13,440 responses. The first repetition of each type was also presented to the 20 listeners without masking noise. In this condition there

were 24 types spoken by 7 speakers for a total of 168 tokens yielding 3,360 responses.

The stimuli were prepared in the following way. First, the stimuli were digitally edited on SoundScope so that all but the consonant and 100 ms. of the vowel were removed. The consonant included the release and following frication and/or aspiration. The 100 ms. vowel duration was determined from the beginning of the periodic structure of the speech wave form. The remainder of the word was truncated at a zero crossing in the wave form 100 ms. inside the vowel. After editing, the stimuli were normalized for intensity. The Root Mean Square (RMS) of the voltage was taken from a 50 ms. window in the vowel. The left edge of the window was at least 3 glottal pulses inside the vowel. Then the whole demisyllable was multiplied by a factor to make the measured RMS equal to 2 volts. In this way the stimuli were normalized for intensity while still preserving the inherent amplitude difference between the consonant and the vowel of the demisyllable.

This was all the preparation done on the stimuli for the no noise condition. In the noise condition, white masking noise was added to stimuli. The signal to noise ratio (S/N) was set at +2dB. This level was determined by a series of pilot experiments which investigated S/Ns from +10dB to -2dB. The S/N of +2dB was selected since the level of error it produced was high, but not as high as random guessing. The voltage level of the noise for a +2dB S/N was determined by solving the following equation for  $x$ .

$$20 \log \frac{2}{x} = 2dB$$
$$x = 1.588$$

Thus, if we plug 1.588 volts into the  $x$  of the equation and take the log of the signal to noise ratio (i.e.,  $\log \frac{2}{1.588}$ ), we get .1. The resultant .1 is then multiplied by 20 to give 2dB. The multiplier 20 is the combination of two factors: 10 and 2. The factor of 10 converts from bel to decibel and the factor of 2 is introduced because voltage is equivalent to sound pressure. The measure of sound pressure is equal to the square of intensity. Thus, when intensity is converted to dB the log of the ratio is multiplied by 10, but when sound pressure is converted to dB the log of the ratio is multiplied by 20 (see Ladefoged (1996:80-87) for a more detailed explanation).

In the next step, white noise of 1.588 volts was added to the stimuli using the capabilities of SoundScope. This was done by adding gaussian noise of the same sampling rate (22,500 samples per second) and same duration as the demisyllables. The resultant stimuli had a signal to noise ratio of +2 dB and the noise and speech signal were equal in length.

The stimuli were then randomized (the stimuli with and without noise were kept in separate groups). The randomized demisyllables were recorded onto normal audio tape in blocks of ten with an interstimulus interval of two seconds. There were 8 seconds between each block of ten. The noise stimuli were also divided into three groups (A, B and C) to be presented to the subjects on a rotational basis.

24 subjects were then presented the stimuli. All subjects were undergraduates in linguistics courses at the University of Texas at Austin and were not paid for their participation. The subjects had no known hearing loss and were all native speakers of American English. The subjects were first presented 120 warm-up stimuli. The signal

to noise level was increased for each of three sets of 40 stimuli from +10dB S/N to +6dB S/N and finally to +2dB S/N. Then the subjects were presented with the three groups of noise stimuli. Group A had 230 stimuli, Group B had 230 stimuli, and Group C had 220 stimuli. The order of the three groups was rotated for each set of listeners. Finally the listeners were presented with the 170 stimuli which had no masking noise.

The subjects were given an answer sheet and instructed to circle one of four possible answers to indicate the consonant they heard. The choices **k ch j g** were printed for each response on the answer sheet. I went over the ‘sound’ of each letter with the subjects before they began. The subjects were told to guess if they were not sure of the answer and to try to respond to as many of the stimuli as possible.

The responses to the no-noise stimuli were scored first and used to evaluate the subject’s performance on the task. To be considered in the results, the subjects had to get better than 90% of the responses correct. This percentage was used since the pilot work indicated that most people got over 90% correct on the no-noise stimuli. Four of the subjects scored below 90%. Their answers were discarded. The remaining 20 subjects are considered in the results. The warm-up section was not scored. It served only to help the subjects get used to the task.

#### 4.3.2. Results

The results from the noise masking perception experiment indicate that [k] is heard as [tʃ] more often than any other confusion. Moreover, there is a vowel effect in the [k]/[tʃ] confusion such that [ki] and [ku] sequences are more often heard as [tʃ] than [ka] sequences are. More than 30% of the [k] plus high vowel sequences were

heard as [tʃ] while only 13% of the [kɑ] combinations were heard as [tʃ]. The next most common confusion is for [dʒ] to be misheard as [tʃ]. Approximately 25% of the [dʒ]s were heard as [tʃ]. The third most common confusion was [g] heard as [dʒ]. Around 15% of the [g] tokens were heard as [dʒ]. More [g]/[dʒ] confusions occurred before the vowel [i].

The results of Experiment 3 clearly support the first prediction that there would be more [k]/[tʃ] confusion than [g]/[dʒ] confusion. About 15% of the [g] tokens were heard as [dʒ], whereas around 26% of the [k] tokens were heard as [tʃ].

The second prediction that there would be more velars heard as palatoalveolars before the high front vowel is partially borne out. In the case of [k], there are many more palatoalveolar identifications before [i] than [ɑ]. However, there are almost as many palatoalveolar identifications before [u] as there are before [i]. These results differ from Experiment 2 in which there was a clear split between [i] and [ɑ u] in terms of palatoalveolar response. Here, in Experiment 3, the high vowels function as a class in opposition to the low vowel. The differing results of the two experiments must be a reflection of the effects of gating versus noise masking. In the truncated condition, durational effects were lost. In the noise-masking condition durational cues are kept and the noise would seem to predispose an affricate response. As will be illustrated in the discussion below, however, durational cues alone cannot account for the difference between the responses for Experiment 2 and 3 since the [kɑ] sequences do not have shorter VOTs than the [ki] and [ku] sequences. The frequency characteristics of the stop consonants must also come into play. The mere fact that there is a consistent vowel asymmetry effect indicates that the listeners are basing their responses on some

acoustic/auditory cue in the stimuli. I suggest that the higher average peak spectral frequency of [ku] as compared to [ka] plays a role in the greater palatoalveolar responses.

The third prediction that faster speech productions of [k] would more often be heard as [tʃ] than citation productions is, however, not supported. The results from an analysis of variance indicate that speech style does not have a significant effect on the number of palatoalveolar responses to [k]. In the case of the voiced velar [g] speech style does have an effect, but in the opposite direction than predicted. I suggest that there is some type of a ceiling effect in this data. The signal-to-noise ratio was quite low (i.e. the noise itself was quite loud). Perhaps the number of incorrect responses was pushed to the limit so that the more subtle effect of speech style was lost. Further research could test this proposal with another perception experiment with higher signal-to-noise ratio (i.e. noise of a lower voltage).

The overall results from the noise stimuli are given in Table 4.3. The stimulus consonant is listed on the top row of the table and the stimulus vowel is listed on the second row. The response consonant is listed in the far left column. The diagonal running from top left to bottom right gives the percent correct identifications. Each consonant vowel sequence received more than 1,000 responses. The raw numbers have been transformed into percentages for presentation here. For example, given the consonant vowel sequence [ki], the listeners responded [k] 43% of the time, [tʃ] 35% of the time, [g] 10% of the time, and [dʒ] 12% of the time.

### Overall Results for Noise Stimuli

Heard	Spoken											
	k			tʃ			g			dʒ		
	i	ɑ	u	i	ɑ	u	i	ɑ	u	i	ɑ	u
k	43%	84%	46%	10%	10%	13%	4%	4%	5%	9%	2%	11%
tʃ	35%	13%	31%	85%	87%	84%	4%	*	3%	28%	23%	22%
g	10%	3%	12%	*	*	*	71%	87%	76%	12%	10%	17%
dʒ	12%	*	11%	5%	3%	3%	21%	9%	16%	51%	65%	50%

Table 4.3

The average correct response for the noise-masked stimuli was 69% correct which is above chance (which would be 25% correct). Notice that the largest percentage of errors is to be found in the cell containing [k] heard as [tʃ] and that there are more [tʃ] responses before [i] and [u] than before [ɑ]. Also note that [g] is most often misheard as [dʒ], especially the [gi] sequence. Another interesting finding is that [dʒ] is often heard as [tʃ].

A repeated measures ANOVA was run on the palatoalveolar responses to the velar stimuli. In this test the percentage of [tʃ] and [dʒ] responses to velar stimuli were combined for each of the 20 listeners. The combination was necessary to allow a comparison between the responses to [k] and [g] in the same design. The ANOVA had

three factors. The first factor, speech style, tested for an effect of citation vs. faster speech on the percentage of palatoalveolar responses. The second factor, voicing, tested for an effect of the voicing of the stimulus consonant (i.e., [k] or [g]) on the percentage of palatoalveolar responses. The third factor, vowel context, tested for a vowel effect on the percentage of palatoalveolar responses. Table 4.4 gives the results of the ANOVA. Note that the factors of speech style, voicing, and vowel context all have significant effects. This tells us that the speech style, voicing, and vowel context all have an effect on the percentage of palatoalveolar responses.

Repeated Measures ANOVA for Percentage Palatoalveolar Response to Velar Stimuli  
 (\*=  $p < .05$ )

<b>Main Effects</b>	<b>Degrees of Freedom</b>	<b>F-Value</b>	<b>P-Value</b>
Style (citation, faster)	1,19	8.68	.008*
Voicing (voiceless, voiced)	1,19	28.84	.000*
Vowel Context ([i] [u] [a])	2,38	32.50	.000*
<b>Interactions</b>			
Style*Voicing	1,19	1.20	.287
Style*Vowel Context	2,38	9.01	.001*
Voicing*Vowel Context	2,38	13.49	.000*
Style*Voicing*Vowel	2,38	6.34	.004*

Table 4.4

The following Table 4.5 gives the answers to the stimuli without noise. Note that the correct response rate (as found in the diagonal from top left to bottom right) is

much higher than in the noise condition. In fact, in all but one case, the percentage of correct answers is higher than 90%. The average percent correct is 96. The highest incorrect percentages are to be found in the cell that holds the [tʃ] responses to [dʒ] stimuli. Also note that the 5% [g] responses to the [ki] stimuli are all from the same token which had an abnormally short VOT.

Overall Results for No-Noise Stimuli

Heard	Spoken											
	k			tʃ			g			dʒ		
	i	a	u	i	a	u	i	a	u	i	a	u
k	95%			1%			1%			*		
		100%			*			1%			*	
			100%			*			1%			*
tʃ	*			96%			*			8%		
		*			96%			*			13%	
			*			98%			*			7%
g	5%			*			98%			2%		
		*			*			99%			*	
			*			*			99%			*
dʒ	*			3%			1%			90%		
		*			4%			*			87%	
			*			2%			*			93%

Table 4.5

In the following sections, I present the results of the noise perception test for each stimulus consonant separately. The results of the citation and faster speech stimuli are also considered separately. Since the ANOVA run on the responses to the velars found the factors of voicing, speech style, and vowel context to have a significant effect,

I ran further statistical tests for [k] and [g] separately. These tests kept the same basic design as the overall ANOVA. The data was just split into two groups ([k] and [g]) using the same factors of speech style and vowel context. I report the findings from these *post hoc* tests in the following sections.

#### 4.3.2.1 [k] stimuli

In Table 4.6, the responses to [k] are listed. This time the responses to citation and faster speech stimuli have been separated. Notice that the overall proportion of [k] to palatoalveolar responses changes little in the two speech styles (it is 2.3 in citation speech and 2.2 in faster speech). The main difference is in the percentage of voiced [g] responses: the number increases from an average of 4% in citation speech to 12% in faster speech. Also notice that, in both speech styles, there are more palatoalveolar responses to [ki] and [ku] than to [kɑ].

### Results for [k] Noise Stimuli

Citation Speech Stimuli				Faster Speech Stimuli			
Spoken				Spoken			
Heard	i	a	u	Heard	i	a	u
k	45%	89%	50%	k	42%	79%	43%
tʃ	37%	10%	37%	tʃ	33%	16%	25%
g	7%	*	4%	g	12%	4%	19%
dʒ	11%	*	9%	dʒ	13%	1%	13%

Table 4.6

A repeated measures ANOVA run on the percentage of palatoalveolar responses to the [k] stimuli revealed no significant effect for speech style. There was, however, a significant effect for vowel context. Table 4.7 gives the results of the ANOVA. The two factors were designed as speech style (citation, faster) and Vowel context ([i] [u] [a]). As in the overall test of velar stimuli, the percentage of [tʃ] and [dʒ] responses to [k] stimuli were combined for each of the 20 listeners.

Repeated Measures ANOVA for Percentage Palatoalveolar Response  
to [k] Stimuli  
(\* =  $p < .05$ )

<b>Main Effects</b>	<b>Degrees of Freedom</b>	<b>F-Value</b>	<b>P-Value</b>
Style (citation, faster)	1,19	.62	.440
Vowel Context ([i] [u] [ɑ])	2,38	39.77	.000*
<b>Interactions</b>			
Style*Vowel Context	2,38	3.90	.029*

Table 4.7

Figure 4.1 presents the palatoalveolar responses to [k] in a way that offers a better inspection of the effects of speech style and vowel context. The height of the bars represents the average number of combined [tʃ] and [dʒ] responses to [k] stimuli. There were a total of 168 [k] stimuli (3 vowel contexts \* 2 speech styles \* 4 repetitions \* 7 speakers) and each of the 168 stimuli were presented to 20 listeners. Since 20 subjects heard each token, there are a total of 20 responses per token. The vertical scale of the graph indicates the average number of responses per token. The scale only goes up to 12 responses to fit the data presented, but 20 responses would be possible since there were 20 listeners. The bars are divided into three groups based on the vowel context. And each vowel context is divided by speech style (clear bars are citation speech and shaded bars are faster speech).

### Average Number of Palatoalveolar Responses to [k] stimuli by Vowel Context and Speech Style

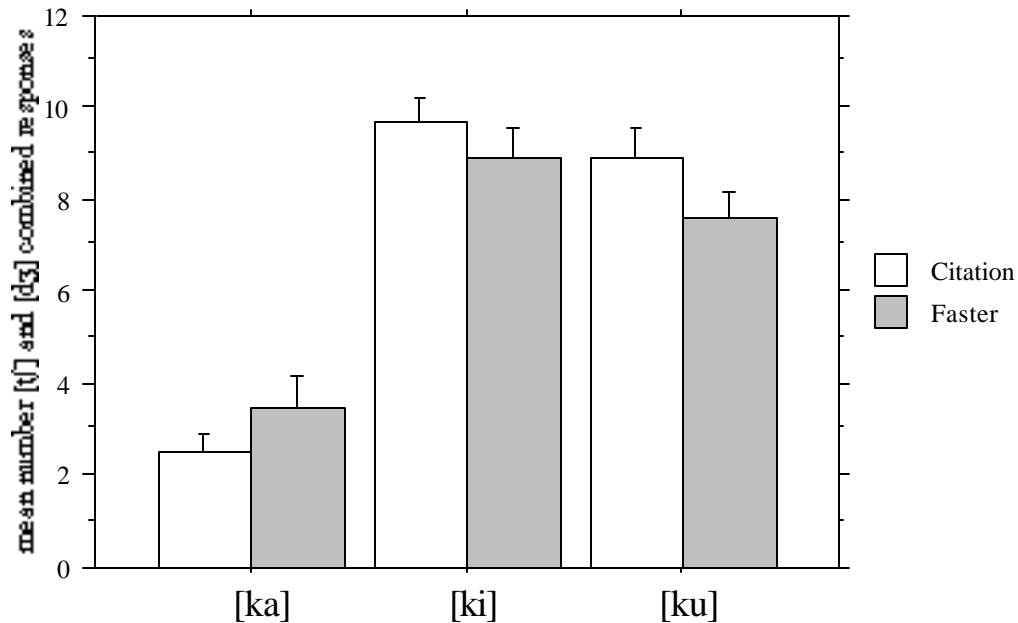


Figure 4.1

Notice that there is very little variation in the average number of palatoalveolar responses by speech style and that there is not a consistent trend across the vowel contexts for more palatoalveolar responses in either speech style. There is, however, a quite obvious trend for the high vowel context to receive more palatoalveolar responses. [ki] receives the most palatoalveolar responses, followed closely by [ku]. The number of [ka] responses is much smaller.

Since the tokens used in the perception experiment are the same as those used in the acoustic study, we can see if there is any relationship between the patterning of the responses and the acoustic attributes of the CV sequences. Here we will examine

the VOT of the [k] tokens as well as the peak spectral frequency of the burst and aspiration of [k].

#### 4.3.2.1.1. VOT of [k] stimuli

First, let us consider the VOT. As discussed in Chapter 3 (§3.4.3), the mean VOT of [k] is 101 ms. in citation speech and 64 ms. in faster speech. The duration of [tʃ] (from release to the onset of the vowel) overlaps with the VOT of [k]. The mean duration of [tʃ] in citation speech is 125 ms. and in faster speech is 91 ms (see Figure 3.31). Figure 4.2 illustrates the average VOT of the [k] stimuli used in Experiment 3. The [k]'s are split by speech style and vowel context. Note that the citation stimuli have longer VOTs across the board. The [kɑ] stimuli have the longest VOT in citation speech, and the [ki] stimuli have the longest VOT in faster speech. The average length of [tʃ] (before all vowel types) is noted with the dotted lines. The average VOT of the citation [k] stimuli fall within the [tʃ] range.

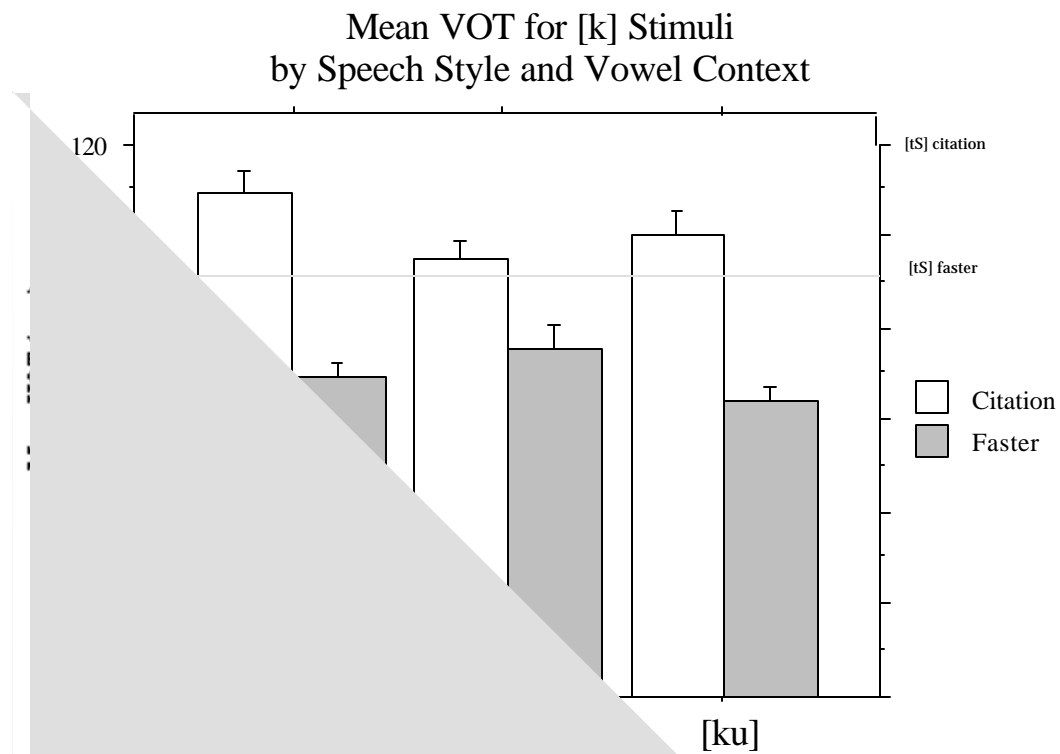


Figure 4.2

Obviously, the VOT of [k] is not the only factor contributing to the number of palatoalveolar responses. If that were the case, we would expect the citation forms of [kɔ] to have the largest average palatoalveolar response. A quick look at Figure 4.1 will show this not to be the case. The [ki] stimuli have the greatest number of palatoalveolar responses. However, if the length of [k]'s VOT plays a role in contributing to [tʃ] responses, we would expect there to be more [tʃ] responses in the VOT range that overlaps with the duration of [tʃ]. If we factor out vowel context by looking at the number of responses as a function of VOT for each vowel context individually, we can investigate the role of VOT in conditioning palatoalveolar responses.

Figures 4.3-4.5 below illustrate the average number of responses for [k] stimuli for a given length of VOT. Since 20 subjects heard each token, there are a total of 20 responses per token. The vertical scale of the graph indicates the average number of responses per token from 0 to 12. The horizontal scale groups the tokens in terms of a 20 ms. VOT range. The columns each represent a certain type of incorrect response. The white columns represent the average number of [dʒ] responses; the lightly shaded columns represent the average number of [g] responses; and the darkly shaded columns represent the average number of [tʃ] responses.

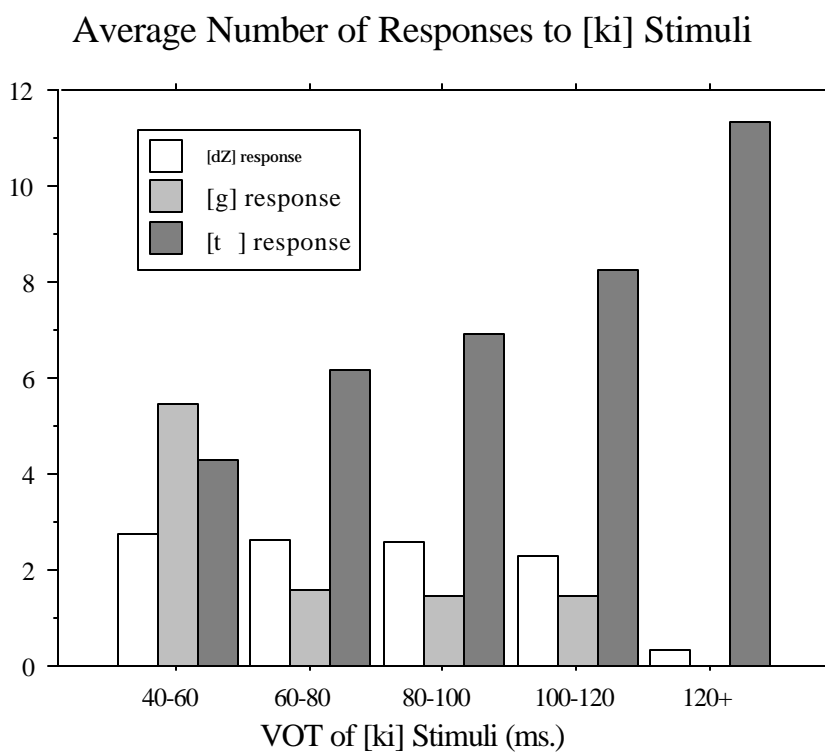


Figure 4.3

First let us consider the responses to the [ki] stimuli. Figure 4.3 shows the average responses to [ki] stimuli. Note that there is a clear trend of increased [tʃ] response as VOT increases. The longest VOTs have an average [tʃ] response of almost 12. This is more than half of the responses. In other words, the [ki] stimuli with the longest VOT are more often heard as [tʃ] than [k]. A linear regression analysis of [tʃ] responses to length of [ki] VOT reveals that this is a significant trend ( $F(1,54)=16.157$ ,  $p=.0002$ ,  $R^2=.23$ ). There is also a clear trend in decrease of [g] response with increase of VOT. With a small VOT of 40-60 ms. the average number of [g] response is around 5 (a quarter of the 20 responses). As the VOT increases, the number of [g] responses drops to less than one. A linear regression analysis which has duration of [k] VOT as an independent variable and average number of [g] responses as the dependent variable returns a significant finding ( $F(1,166)=26.632$ ,  $p < .0001$ ,  $R^2=.14$ ).

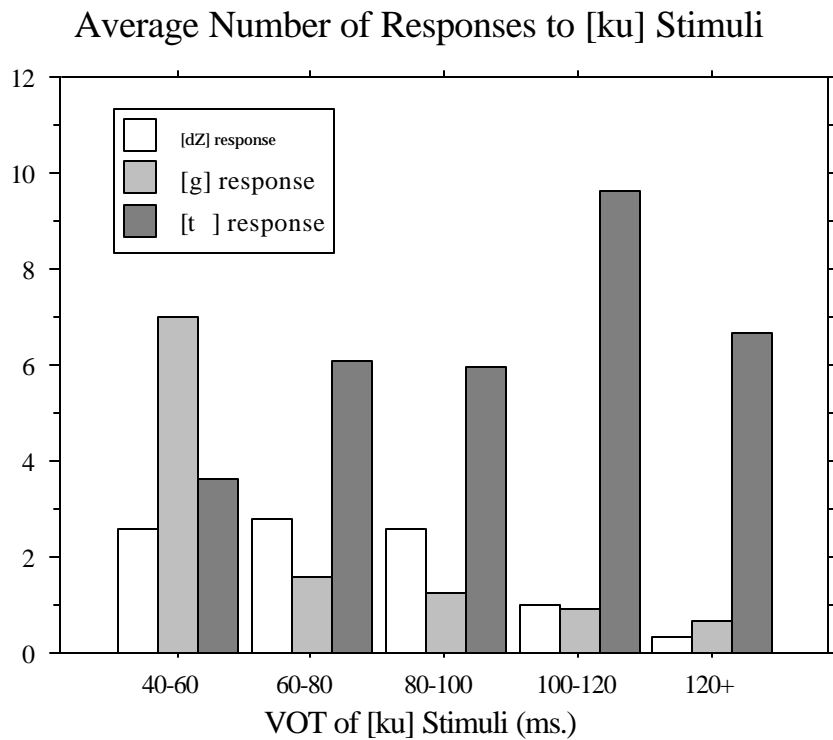


Figure 4.4

The responses to [ku] stimuli do not provide such clear results. Note that there are more [tʃ] responses above 60 ms., but that there is not a linear trend of increased [tʃ] response to longer VOT stimuli. However, as in the case of the [ki] stimuli, the number of [g] responses clearly decreases with shorter VOT stimuli. In the 40-60 ms. range, there is an average of 7 [g] responses out of 20. This number drops dramatically into the longer VOT ranges.

### Average Number of Responses to [kɑ] Stimuli

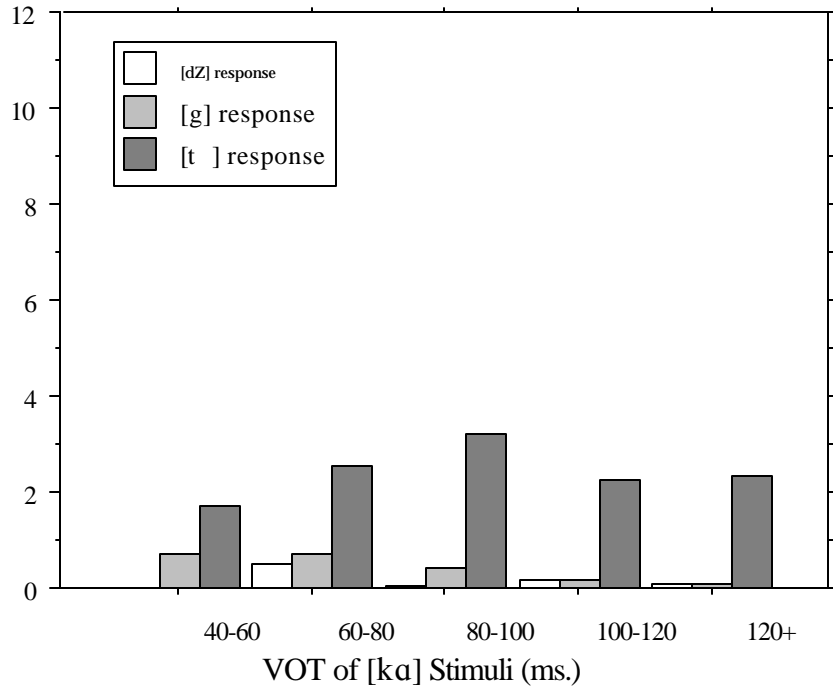


Figure 4.5

Finally, the [kɑ] stimuli have a much smaller overall incorrect response rate. The most common misidentification is [tʃ], but the average number of [tʃ] responses never goes above 4. The trend towards more [g] responses in the shorter VOT ranges is weakly echoed in the [kɑ] data.

To sum up the discussion of [k]’s VOT, the length of VOT has a differential effect on the number of [tʃ] responses in the three vowel contexts. The [ki] stimuli showed a clear trend to increase the number of [tʃ] responses as the VOT increased. The [ku] stimuli showed a weaker version of this trend, and the [kɑ] stimuli did not show this trend. This suggests that the length of VOT influences the responses only in

conjunction with some other factor. Perhaps the peak spectral frequency of the [k] needs to be in the [t ] range for the VOT effects to kick in. Or perhaps some other factor not investigated here plays a role. We will come back to the interaction of VOT and peak spectral frequency at the end of the next section.

#### 4.3.2.1.2. Peak Spectral Frequency of [k] stimuli

Now let us consider the relationship of responses to [k] stimuli and the peak spectral frequency of the [k] burst and aspiration. If the peak spectral frequency of [k] plays a role in number of [tʃ] responses in the perception test and by inference to the [k] to [tʃ] sound change, we would expect more [tʃ] responses when the peak spectral frequency of [k] overlaps that of [tʃ]. For male speakers, the average peak spectral frequency of [tʃ] ranges from 3800-4000 Hz. For female speakers, the peak spectral frequency for [tʃ] is around 4000-4300 Hz. Figure 4.6 illustrates the average peak spectral frequency of the [k] stimuli used in Experiment 3. The [k]'s are split by speech style and vowel context. Note that the faster stimuli have higher peak spectral frequencies across the board. The [ki] stimuli have the highest frequencies, then the [ku] stimuli, and finally the [kɑ] stimuli. Higher peak spectral frequency of the [ku] stimuli (as compared to the [kɑ] stimuli) is probably a reflection of the relatively fronted production of [u] in Texas English. The average peak spectral frequency of [tʃ] (before all vowel types) is noted with the dotted lines. The average peak spectral frequency of the faster [ki] stimuli fall within the [tʃ] range.

### Mean Peak Spectral Frequency for [k] Stimuli by Speech Style and Vowel Context

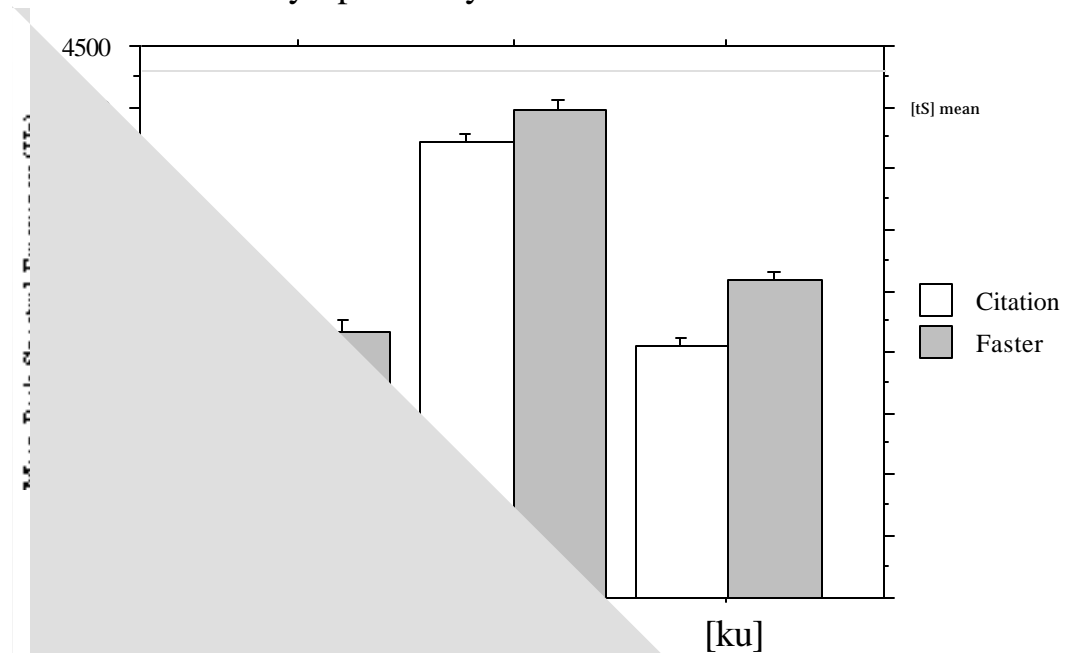


Figure 4.6

Based on Figure 4.6, we would expect the [ki] stimuli to be most confusable with [tʃ]. The [ki] stimuli receive the most palatoalveolar responses, but [ku] comes in a close second as can be seen in figure 4.1. We do find, however, that the most palatoalveolar responses are found in roughly the range of the average [tʃ] peak spectral frequency. Consider Figure 4.7 below. The average number of responses (again, out of 20) to [k] are represented along the vertical axis. The peak spectral frequency is indicated along the horizontal axis. Note that the greatest average [tʃ] response is to be found from 3000-4500 Hz. Responses taper off before and after this region. Here we have evidence that the acoustic similarities between [k] and [tʃ] which we found in the

last chapter do play a role in perception. Those tokens of [k] which have a peak spectral frequency within the range of [tʃ] are more often confused with [tʃ].

Average Number of Responses to [k] Stimuli

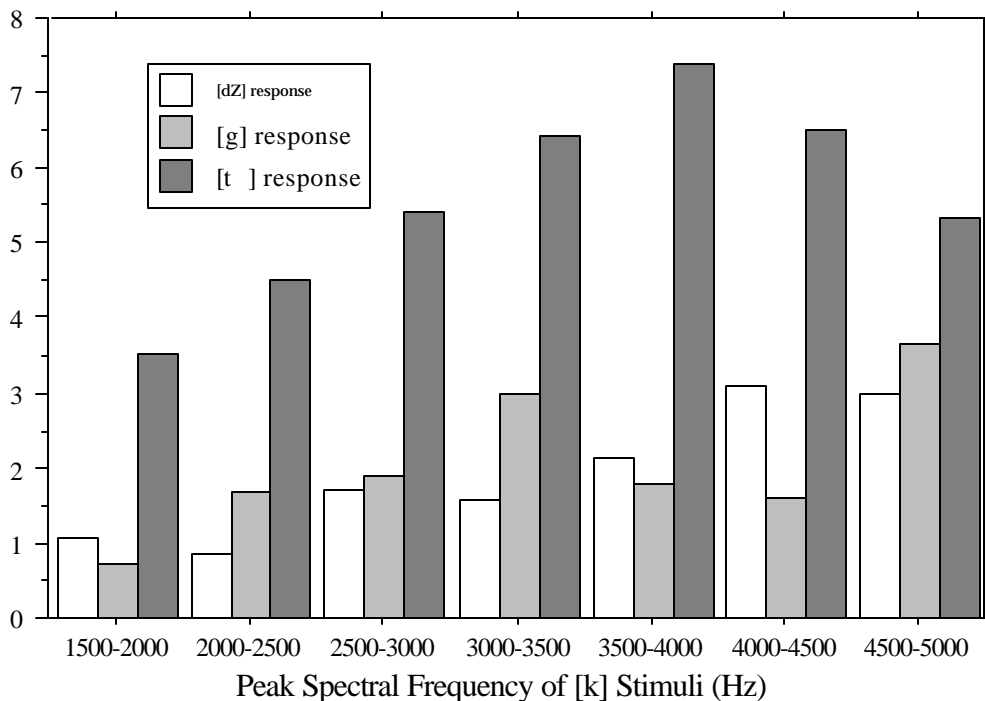


Figure 4.7

Now, I return to the interaction of the factors of VOT and peak spectral frequency. We have seen that the number of [tʃ] responses to [k] stimuli increases as VOT increases. This trend was the most clear in the [i] vowel context. I suggested at the end of the last section that the peak spectral frequency of [k] needed to reach some critical level for the VOT effect on [tʃ] responses to be readily noticed. Figure 4.8 provides a graphical representation of the interaction of VOT and peak spectral frequency. The x-axis lists the peak spectral frequency of the [k] stimuli. The length of

the bars represents the average number of [tʃ] responses as defined by the numbers on the y-axis. For each frequency range, the responses have been split by VOT.

Mean Number of [tʃ] Responses to [k] Stimuli  
by Peak Spectral Frequency and VOT of [k]

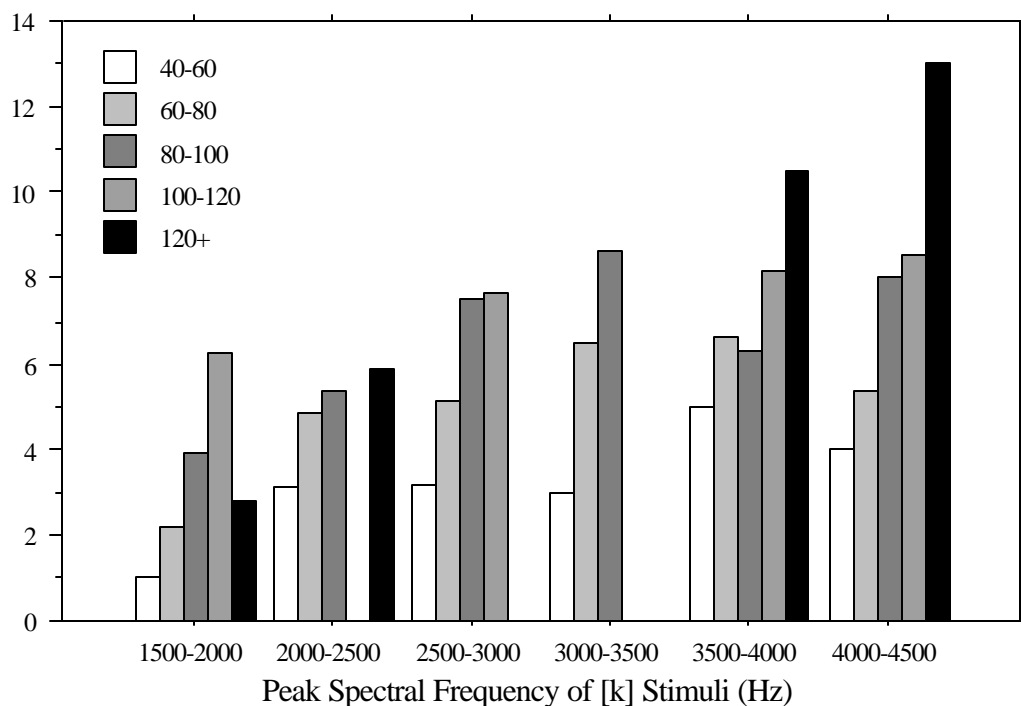


Figure 4.8

First, note that as the burst frequency of [k] increases the number of [tʃ] responses increase. Second, note that, in general, within each frequency range, the number of [tʃ] responses increases as the VOT increases. The 1500-2000 Hz range is the only exception to the trend. For example, in the 2000-2500 Hz range the mean number of [tʃ] responses increases from 3 to 6 over the range of VOTs. In the 4000-4500 Hz range the mean number of [tʃ] responses increases from 4 to 13 over the range of

VOTs. So, the effect of VOT on [tʃ] response is unclear for the lowest frequency range. The average number of [tʃ] responses increases by 3 for the 2000-2500 Hz range, by 5 for the 2300-3000 Hz range, by 6 for the 3500-4000 Hz range, by 6 for the 4000-4500 Hz range, and by 9 for the 4500-5000 Hz range. Thus, we can see that as the peak spectral frequency of [k] increases, the VOT plays a larger role in determining the number of [tʃ] responses.

#### 4.3.2.2. [g] stimuli

Next I present the results of the [g] stimuli. Table 4.8 below gives the percentage of [k], [tʃ], [g] and [dʒ] responses to [g] stimuli in the two speech styles. Notice that most of the errors lie in hearing [g] as [dʒ] and that [g] is most often misheard as [dʒ] before [i]. Contrary to prediction, the biggest change from identifications for citation speech to identifications for faster speech is a decrease in the number of [gi] sequences heard as [dʒ].

## Results for [g] Noise Stimuli

Citation Speech Stimuli				Faster Speech Stimuli			
Spoken				Spoken			
Heard	i	a	u	Heard	i	a	u
k	5%	*	3%	k	3%	7%	6%
tʃ	6%	*	3%	tʃ	2%	2%	3%
g	64%	91%	77%	g	79%	82%	76%
dʒ	25%	8%	17%	dʒ	16%	9%	15%

Table 4.8

A repeated measures ANOVA run on the percentage of palatoalveolar responses to the [g] stimuli revealed a significant effect for speech style and for vowel context. Table 4.9 gives the results of the ANOVA. The two factors were speech style (citation, faster) and Vowel context ([i] [u] [a]). The percentage of [tʃ] and [dʒ] responses to [g] stimuli were combined for each of the 20 listeners.

Repeated Measures ANOVA for Percentage Palatoalveolar Response  
to [g] Stimuli  
(\* =  $p < .05$ )

<b>Main Effects</b>	<b>Degrees of Freedom</b>	<b>F-Value</b>	<b>P-Value</b>
Style (citation, faster)	1,19	17.52	.001*
Vowel Context ([i] [u] [ɑ])	2,38	9.20	.001*
<b>Interactions</b>			
Style*Vowel Context	2,38	13.82	.000*

Table 4.9

Figure 4.9 presents the palatoalveolar responses to [g]. The height of the bars represents the average number of combined [tʃ] and [dʒ] responses to [g] stimuli. As in the case of [k], there were a total of 168 [g] stimuli and each of the 168 stimuli were presented to 20 listeners. Since 20 subjects heard each token, there are a total of 20 responses per token. The vertical scale of the graph indicates the average number of responses per token. The bars are divided into three groups based on the vowel context. Each vowel context is divided by speech style.

### Average Number of Palatoalveolar Responses to [g] stimuli by Vowel Context and Speech Style

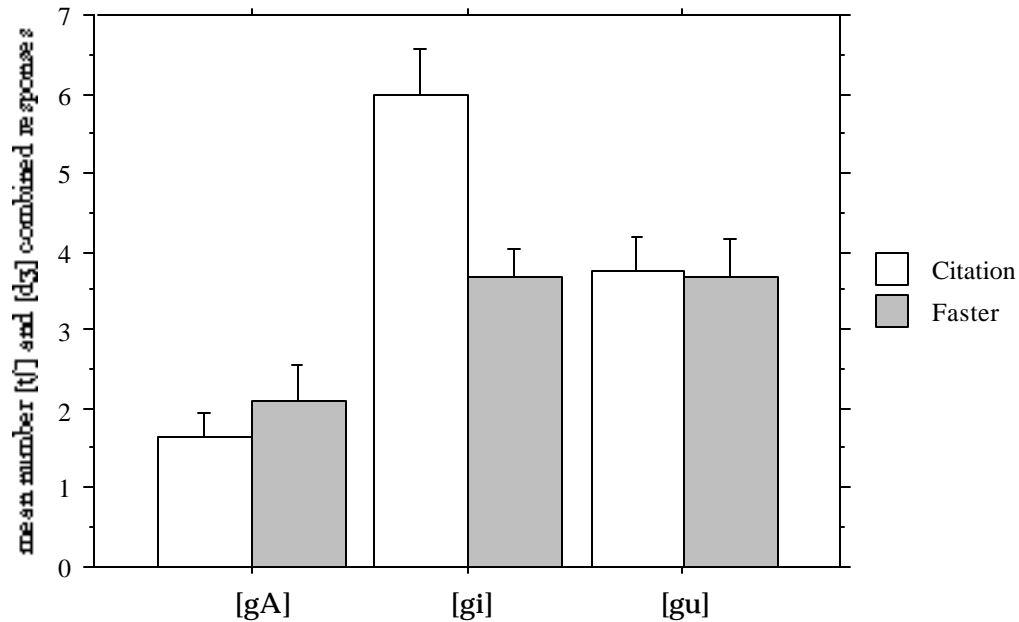


Figure 4.9

Note that the largest palatoalveolar response is found for the [gi] citation stimuli, with an average of 6 palatoalveolar responses out of the possible 20. The [gi] faster and [gu] (citation and faster) stimuli have an average palatoalveolar response rate of 4. The [gA] stimuli have a lower response rate around 2.

Analysis of the acoustic attributes of the [g] stimuli reveal that length of VOT and peak spectral frequency have an effect of the number of [dʒ] responses.

#### 4.3.2.2.1. VOT of [g] stimuli

The VOT of the [g] stimuli is well below that of palatoalveolars. The average duration of [dʒ] is 53 ms. in faster speech and 71 ms. in citation speech. As can be

seen in Figure 4.10, the [gi] tokens in faster speech have the longest VOTs at around 33 ms. Comparison with Figure 4.9 reveals that the tokens with the longest VOT ([gi] citation) are also the tokens with the highest palatoalveolar response.

Mean VOT for [g] Stimuli  
by Speech Style and Vowel Context

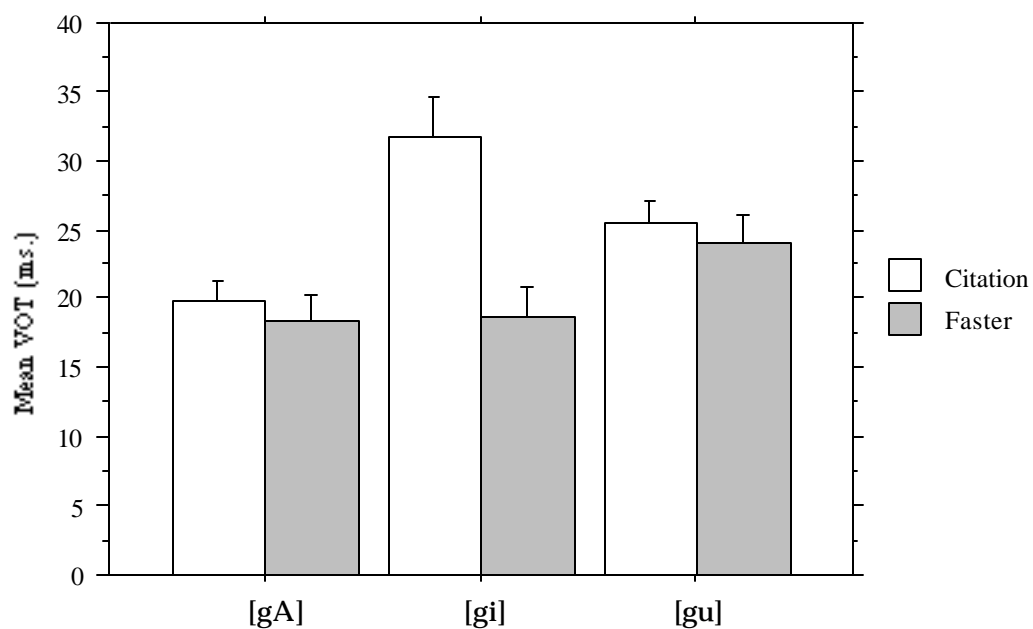


Figure 4.10

Figure 4.11 below shows that stimuli with VOT over 30 ms. have an average [dʒ] response rate of over 4. Those stimuli with a VOT of less than 30 ms. have an average [dʒ] response rate of 3 or less. The difference is not large but a regression analysis shows that it is significant ( $F(1,166)=13.211$ ,  $p=.0004$ ,  $R^2=.074$ )

### Average Number of Responses to [g] Stimuli

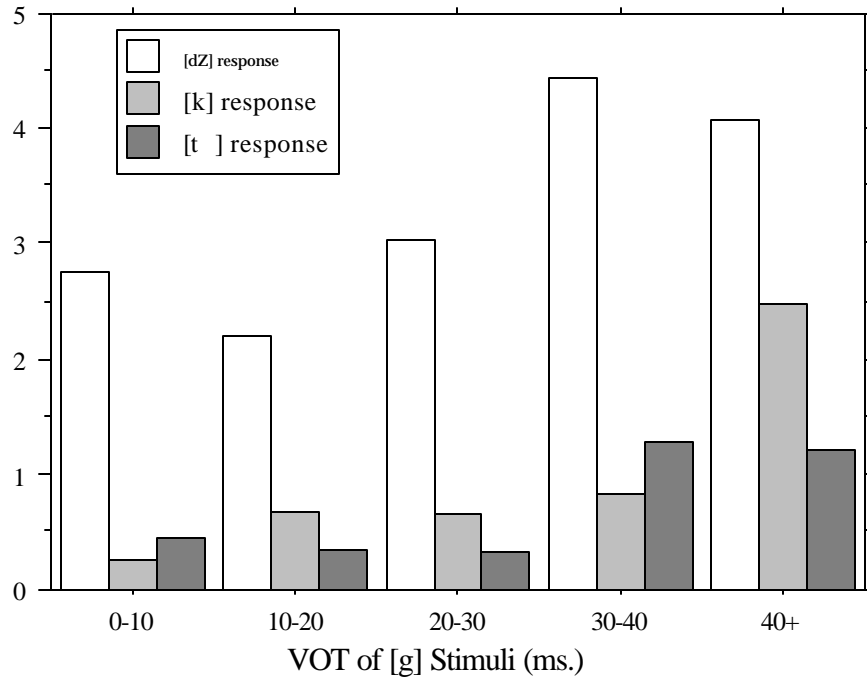


Figure 4.11

#### 4.3.2.2.2. Peak Spectral Frequency of [g] stimuli

The average peak spectral frequency for the [g] stimuli in general falls below the mean for [dʒ]. The mean peak frequency for [dʒ] is from about 3800-4000 for males and from 4000-4500 for females. Figure 4.12 below shows that the [gi] stimuli come the closest to the [dʒ]. Again, this is reflected in the number of palatoalveolar responses displayed in 4.8: the [gi] stimuli have more palatoalveolar responses.

### Mean Peak Spectral Frequency for [g] Stimuli by Speech Style and Vowel Context

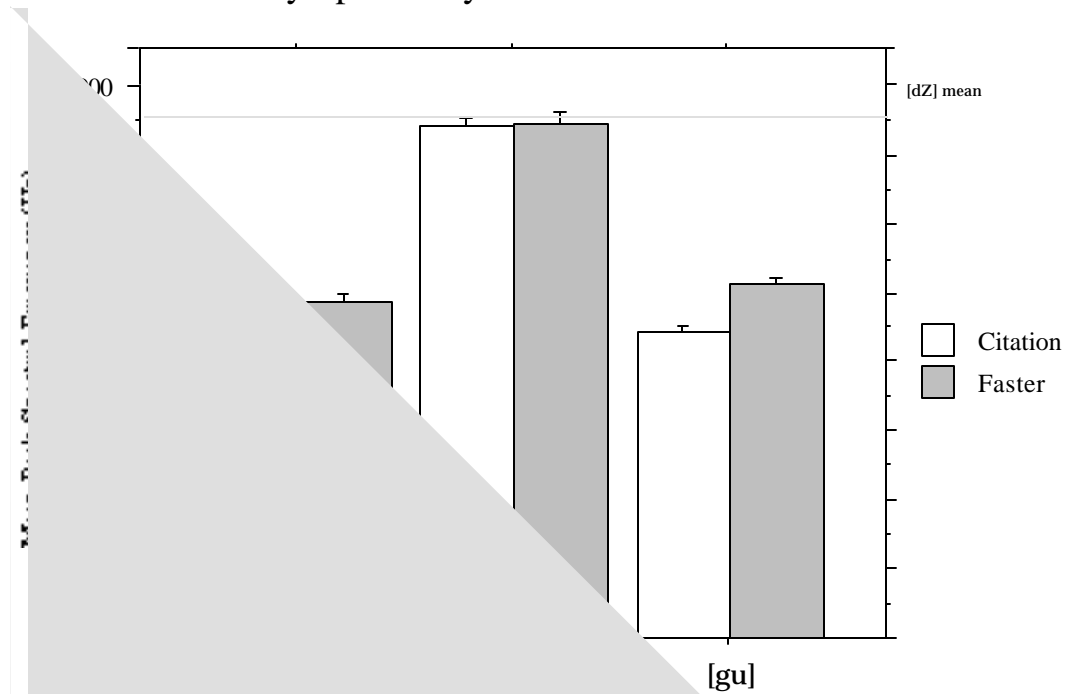


Figure 4.12

The peak spectral frequency of the [g] stimuli also has an effect on the number of [dʒ] responses. As Figure 4.13 indicates, the higher the peak spectral frequency of [g] the greater the average [dʒ] response is. This relationship is significant ( $F(1,157)=17.72, p < .001, R^2=.102$ )

### Average Number of Responses to [g] Stimuli

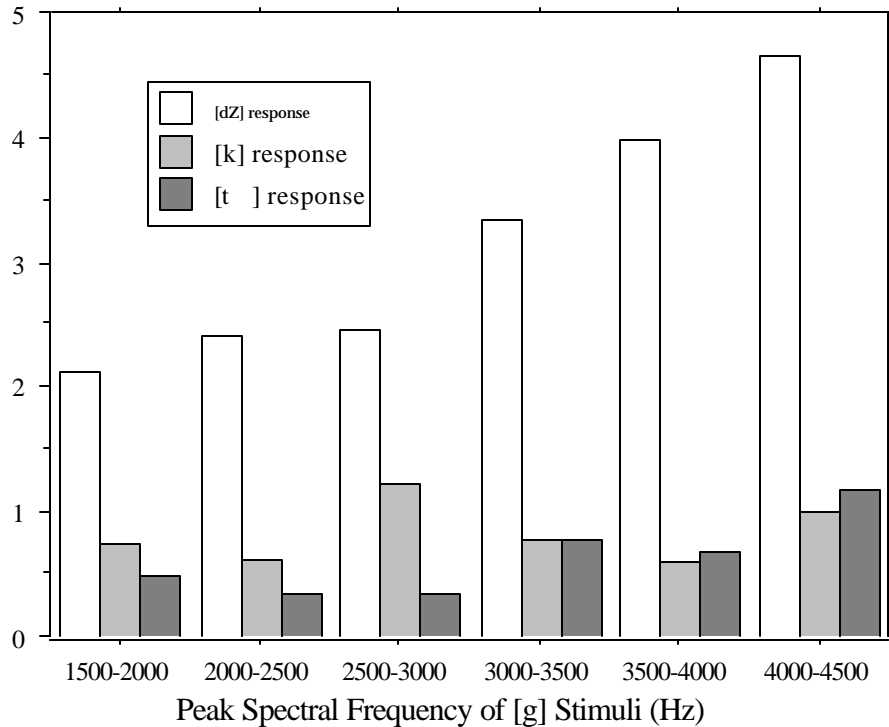


Figure 4.13

#### 4.3.2.2.3. Further Investigation into [g] stimuli: Experiment 3.1

As mentioned in Chapter 2, [g]’s often palatalize to [j]. Since [j] was not one of the options in the forced choice perception experiment, I wanted to see if [g] was more often heard as [dʒ] or as [j]. In Experiment 3.1, I had 5 expert transcribers in the UT Linguistics Department listen to all the [g] stimuli without noise and give a narrow transcription. All the transcribers had extensive experience in fieldwork and training in phonetic transcription. The stimuli were played to the transcribers three times at a 1 second interval. The transcribers then had 5 seconds to transcribe the segment. There

were 168 [g] tokens heard by 5 listeners for a total of 840 responses. The transcribers were told to be as detailed as they could in the time allowed.

Interestingly enough, none of the transcribers heard the [g]s as [dʒ]. They did, however, hear some tokens as [j]. Table 4.10 gives the percentages of the most common transcriptions (other than [g] or [k]). Some of the transcribers also heard [k]'s. These are not included in the chart. Note that only [gi] sequences were transcribed as [j] and that only the faster speech tokens were heard as [j]. Note also that the [gi] faster speech tokens were transcribed as something other than a plain [g] the most often. They were transcribed as [j] 5% of the time, as [gʲ] or [kʲ] 4% of the time, as [ʔ] 4% of the time, and the consonant was not heard at all 3% of the time. This finding parallels the observation that [g] is often palatalized to [j]. The results also support the proposal that faster speech tokens are more likely to be heard as palatal segments.

Transcription of No-Noise [g] Stimuli

Transcription	[g] Citation			[g] Faster		
	i	a	u	i	a	u
[j]	*	*	*	5%	*	*
[gʲ]/[kʲ]	8%	*	*	4%	*	3%
[ʔ]	*	*	*	4%	*	*
no consonant heard	*	*	*	3%	*	*

Table 4.10

The results from the transcription are different from the results of the forced choice perception experiment. Of course the task was different (transcription vs.

circling the best answer) and the stimuli were themselves different (no-noise vs. noise). In the forced choice experiment discussed above, there were more [dʒ] responses in citation speech than in faster speech. In the transcription task there were more [j] responses in faster than in citation speech. One thing the results have in common, however, is that the [gi] sequences are the most likely to be heard as something other than [g]. In some way [g] before [i] is less stable perceptually than [g] before [u] or [ɑ].

#### 4.3.2.3. [dʒ] stimuli

Now let us consider the responses to the [dʒ] stimuli. Table 4.11 gives the percentage of [k], [tʃ], [g], and [dʒ] responses. The results of the citation speech stimuli and the faster speech stimuli have again been separated. Note first that [dʒ] is most often misheard as [tʃ]. The second most common incorrect response is [g]. And the faster speech tokens are more often heard as [k] than the citation speech tokens.

## Results for [dʒ] Noise Stimuli

Citation Speech Stimuli				Faster Speech Stimuli			
Spoken				Spoken			
Heard	dʒ			Heard	dʒ		
	i	a	u		i	a	u
k	7%	*	8%	k	11%	4%	14%
tʃ	30%	15%	22%	tʃ	27%	30%	23%
g	11%	6%	12%	g	13%	14%	21%
dʒ	52%	79%	58%	dʒ	49%	52%	42%

Table 4.11

The duration of the [dʒ] stimuli has an effect on the type of response. As indicated by Figure 4.14, as the duration increases, the average number of [tʃ] responses increases. A linear regression analysis shows that this is a significant effect ( $F(1,166)=67.5$ ,  $p < .0001$ ,  $R^2=.289$ ). Another trend illustrated in Figure 4.14 is that as the duration of [dʒ] decreases, the frequency of stop responses increases. This trend is especially clear with the [g] stimuli. A regression analysis of the average number of [g] responses on the duration of [dʒ] proves significant ( $F(1,166)=39.295$ ,  $p < .0001$ ,  $R^2=.191$ ).

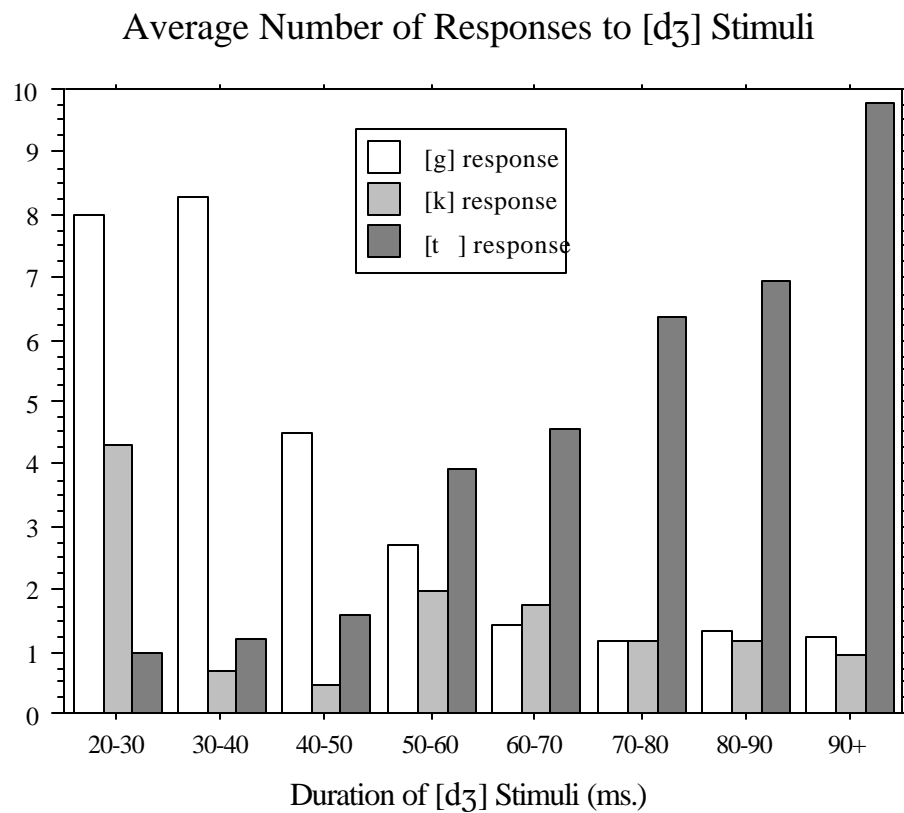


Figure 4.14

#### 4.3.2.4. [tʃ] stimuli

Finally, let us consider the results of the [tʃ] stimuli. Table 4.12 gives the responses to [tʃ]. Again, the responses to the citation and faster speech stimuli have been separated. In general, the [tʃ] stimuli are identified correctly at a greater percentage than the [dʒ] stimuli. Note that there are very few confusions with [dʒ]. The most common incorrect response is [k]. [k] responses also increase in faster speech.

## Results for [tʃ] Noise Stimuli

Citation Speech Stimuli				Faster Speech Stimuli			
Spoken				Spoken			
Heard	i	tʃ		Heard	i	tʃ	
		a	u			a	u
k	8%	4%	15%	k	12%	16%	10%
tʃ	88%	95%	81%	tʃ	80%	79%	87%
g	*	*	1%	g	1%	1%	*
dʒ	4%	*	3%	dʒ	7%	4%	3%

Table 12

The [tʃ] stimuli with the shortest duration are more often heard as [k]. Figure 4.15 shows the average number of responses to [tʃ] as a function of the stimulus duration. Note that the stimuli from 40-80 ms. are heard as [k] have an average response rate of almost 5. Those stimuli over 80 ms. in duration have an average response rate of below 2. This trend is significant as illustrated by a regression analysis ( $F(1,166)=7.998$ ,  $p=.0009$ ,  $R^2=.064$ ).

### Average Number of Responses to [tʃ] Stimuli

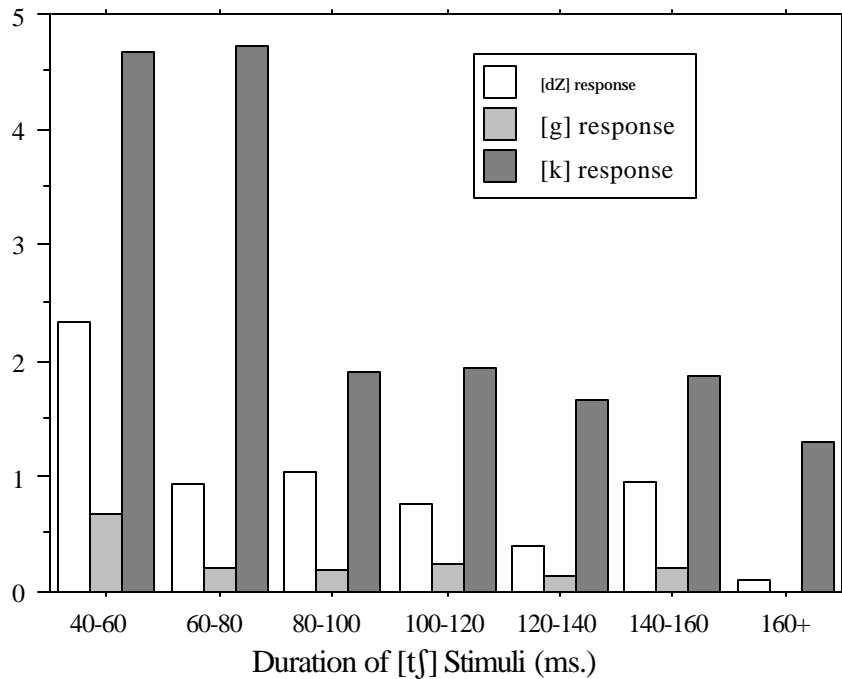


Figure 4.15

#### 4.4. Summary

The experiments discussed here support the hypothesis that [k] is perceptually confusable with [tʃ] before front vowels. The results also agree with the typological evidence that [k] to [tʃ] palatalization is more common than [g] to [dʒ] palatalization. Evidence was also presented that faster speech [g]'s are heard as [j], a common result of [g] palatalization. Table 4.13 summarizes the results of the perception experiments

### Results of Perception Experiments

Noise Masking	Stimulus duration	Results
<b>Experiment 2:</b> stimuli [k tʃ] + [i a u] faster speech		
no	all consonant	98-100% correct ID's
no	30 ms. of consonant	[ki] heard as [tʃ], else 98-100% correct ID's
<b>Experiment 3:</b> stimuli [k tʃ g dʒ] + [i a u] citation and faster speech		
no	all consonant + 100 ms. of vowel	96% average correct ID's
+2 dB S/N	all consonant + 100 ms. of vowel	[ki], [ku] heard as [tʃ], [dʒ] heard as [tʃ], [gi] heard as [dʒ]
<b>Experiment 3.1:</b> stimuli [g] + [i a u] citation and faster speech		
no	all consonant + 100 ms. of vowel	Faster [gi] transcribed as [j gʲ ?] or not heard

Table 4.13

The prediction that faster speech tokens of [k] would be heard as [tʃ] more often than citation speech tokens of [k] did not receive support. This is not surprising when we consider the nature of the noise-masked stimuli. I suggest that the noise had a greater effect on the responses than the speech style. The speech signal was highly

degraded by the masking noise. In essence, all the noise stimuli were hypo-forms. The gross effect of the noise seems to have obviated the faster/citation effect. There was, however, a significant difference in the responses to faster and citation forms for the [g] stimuli. In general, there were more voicing confusions in faster speech. These voicing confusions were mostly due to the duration of the consonant stimuli. Duration is one effect not masked by noise.

The results of the perception experiment complement the findings of the acoustic investigation. The consonants [k] and [tʃ] were found to be similar acoustically, especially before a high front vowel. In the perception experiments, [k] was often heard as [tʃ] when it was before a high front vowel. In Experiment 2, [ki] sequences were heard as [tʃ] about half the time while the [kɑ] and [ku] sequences were heard correctly over 90% of the time. In Experiment 3, the [ki] sequence was the most often heard as [tʃ]. However, the [ku] sequence was close behind. The noise masking in this experiment predisposed an affricate response. The overall high number of palatoalveolar responses to the velars must be attributed, at least in part, to the noisy conditions. It seems that the acoustic/perceptual properties of [ka] were the most clearly non-affricate-like. Perhaps the slightly higher peak spectral frequency of the consonant and the fronted production of the vowel in the [ku] tokens lended itself to an affricate identification.

It is interesting to note that [tʃ] was not heard as [k] nearly so often. In fact, [tʃ] was heard correctly the most often of all the consonants. Other perception experiments including affricates have also found that they are the least often confused with other places and manners of articulation (e.g., Ahmed and Agrawl 1969). The

perception results are consistent with the sound change typology: a [k] becoming [tʃ] is quite common, whereas a [tʃ] becoming [k] is quite rare.

To sum up, the results of the perception experiments parallel the sound change of palatalization in several ways. First, [k] is heard as [tʃ] most often before high front vowels. Second, [k] is heard as [tʃ] more often than [g] is heard as [dʒ] and [g] can be heard as [j]. Thirdly, [tʃ] is rarely confused with [k]. These parallels argue strongly for an account of velar palatalization which involves a perceptual component.