Original Paper

Phonetica 1998;55:18–52

Received: May 18, 1997
Accepted: September 25, 1997

The Role of Perception in the Sound Change of Velar Palatalization

Susan Guignard Guion

The University of Texas at Austin, Department of Linguistics, Austin, Tex., USA

Abstract

Voiceless velar stops may become palatoalveolar affricates before front vowels. This sound change is not only one of the most common types of palatalization, but is a very common sound change in the world's languages. Nevertheless, we do not have an adequate understanding of how this sound change takes place. Three experiments reported here test the hypothesis that velar palatalization is the result of listeners' on-line perceptual reanalysis of fast rate speech. It is shown that velars before front vowels are both acoustically and perceptually similar to palatoalveolars. This supports the proposal that velar palatalization is perceptually conditioned.

Introduction

In the search for conditioning factors in sound change, many types of explanations have been investigated, including systemic, lexical, articulatory, and perceptual factors. The work presented here explores a perceptual explanation for a common change that has occurred historically in many languages. The focus is on the sound change of velar palatalization in which a voiceless velar stop becomes a palatoalveolar affricate. The most common environment for this sound change is before front vowels and palatal glides. The work presented here will concentrate on the front vowel environment. Velar palatalization before front vowels is generally considered to be not only one of the most common types of palatalization, but indeed a very common sound change in the world's languages [Bloomfield, 1933; Chen, 1973; Bhat, 1978; Hock, 1991]. Traditionally, the sound change from [k] to [tʃ] is considered to be a gradual articulatory change. There is not, however, a consensus view on how such an articulatory change might be accomplished. Detailed descriptions of palatalization are few and the accounts vary considerably.

One view proposes that a fronted [k] can develop a fricative release which precipitates a shift in articulation to a more anterior position [Grammont, 1933, p. 214; Hock, 1991, pp. 73–77]. This proposal does not, however, motivate a change from affricated palatal to palatoalveolar. There is no articulatory reason why the sound change could not result in a palatal place of articulation. However, this is a typologically rare outcome of velar palatalization [Bhat, 1978].

KARGER

Fax +41 61 306 12 34
E-Mail karger@karger.ch
www.karger.com

0031–8388/98/0552–0018
$15.00/0
Article accessible online at:
http://BioMedNet.com/karger

Susan G. Guion, University of Alabama at Birmingham
Biocommunication Research Laboratory
VH 503, Birmingham, AL 35294-0019 (USA)
Tel. +205 934 3644, Fax +205 934 7420
E-Mail guionsg@uab.edu

Another view proposes that a fronted velar acquires a palatal offglide and then is shifted to a coronal place of articulation due to the narrow shape of the vocal tract [Anttila, 1989, pp. 72–73]. It remains unclear, however, why a narrowed vocal tract would give rise to fronting. A problem with these accounts is the assumption that palatalization is *purely* an articulatorily motivated change. The two accounts just mentioned begin with a fronted velar, but diverge from there. While it might be possible to imagine a purely articulatory account in which the tongue begins at a velar articulation and gradually creeps up to a palatoalveolar place of articulation, changing the part of the tongue used for the articulation (from dorsum to blade) as well as acquiring a fricative release along the way, the actual fleshing out of this account is problematic.

Ohala [1986, 1989, 1992] has proposed an alternative explanation for velar palatalization. He suggests that there is a *perceptual* motivation for the palatalization of [k] to [tʃ] before front vowels. Ohala [1992, p. 320] argues against an articulatory explanation of palatalization. First, he notes that velar palatalization results in a place of articulation that is further forward than the conditioning environment ([j] or [i]). Undershoot but not overshoot is common in articulatory assimilation. Second, he notes that the active consonant articulator changes as a result of the sound change and that the two articulators are too different to be due to articulatory assimilation. Ohala [1989, pp. 183–185] also mentions the high confusion rate of [ki] for [ti] in a perceptual study by Winitz et al. [1972] and suggests that the sound change [k] > [tʃ] can occur if listeners fail to factor out the effects of the [i] context on realizations of velar stops.

Ohala's proposal is investigated here using an experimental methodology. It is proposed here that traditional explanations fail to motivate each step in the palatalization process on purely articulatory grounds due to the fact that important aspects of velar palatalization are perceptually motivated. Looking to the laboratory for explanations of sound change is now well accepted [Ohala, 1971, 1974, 1981, 1989, 1993]. By assuming a phonetic basis for sound change, as well as assuming that the phonetic variation found in speech production today parallels the variation in the past, laboratory investigation into the origins of sound change can be undertaken on any living language. In Ohala's words,

> ... the ultimate check on any hypothesis about the cause of a particular sound change is to test the hypothesis in the laboratory. If particular sound changes are posited to have a phonetic basis then one should be able to duplicate the conditions under which they occurred historically and find experimental subjects producing 'mini' sound changes that parallel them. It is because of the posited phonetic character of sound change that a laboratory study is possible: were the initiation caused by grammatical and cultural factors, this would be more difficult or perhaps impossible [Ohala, 1993, p. 261].

Of course, the more common and widely attested a sound change is, the more valid an experimental approach will be. As Ohala [1993, p. 238] notes, common sound changes attested independently in substantially the same form in unrelated languages are likely to arise from language universal factors such as the physics and physiology of the vocal tract and the nature of the human perceptual system. All of this makes velar palatalization an optimal choice for a laboratory investigation. It is a common sound change, and is attested in many unrelated languages. Consider the examples in table 1 (which summarizes Guion [1996, pp. 4–22]). The examples of velar palatalization are from a variety of language families, demonstrating the widespread occurrence of this change.

Table 1. Examples of velar palatalization

Language	Change	Environment	Source
Slavic (1st palatalization)	k > tʃ g > ʒ x > ʃ	{j, ĭ, i, e, ɛ, ẽ}	Vaillant [1950], Kiparsky [1963], Shevelov [1965], Matthews [1967], Lunt [1974], Schenker [1993]
Slavic (2nd palatalization)	k > tʃ d > dʒ x > s	{i, ɛ}	Vaillant [1950], Kiparsky [1963], Shevelov [1965], Matthews [1967], Lunt [1974], Schenker [1993]
Indo-Iranian	k > tʃ g > dʒ g^h > $dʒ^h$	{i, e}	Mayrhofer [1972], Hoffmann [1982]
Cowlitz Salish	k > tʃ k' > tʃ' x > ʃ	i	Kinkade [1973]
Bantu	k > ts g > dʒ	j (+{i, e})	Guthrie [1967–1970], Hyman and Moxley [1996]
Bantu	k > tʃ g > z̤	i̟	Guthrie [1967–1970], Myers [1992–1994]
Old to Middle English	k̟ > tʃ ɣ̟ > j k̟k̟ > tʃ g˔g˔ > dʒ	{æ ǣ e ē i ɪ}	Emerson [1903], Campbell [1959], Hogg [1979]
Mam (Mayan)	k > tʃ, tʃ̭̊, c	{i, e}	England [1990]
Old to Middle Chinese	k > tɕ k^h > $tɕ^h$ g > dʑ (ʐ) x > ɕ	j + {i, e}	Cheng [1968], Pulleyblank [1984], Baxter [1992]

Examination of numerous cases of velar palatalization reveals a voicing asymmetry. Voiceless velars are more likely to palatalize than voiced velars. In Bhat's [1978] collection of examples of velar palatalization conditioned by a following front vowel or glide, about 60% of the cases involved both the voiced and voiceless velars and 40% of the cases involved *only* voiceless velars. There were no examples, however, of a voiced velar palatalizing to the exclusion of the voiceless velar. Also note that in the examples given in table 1, the voiceless velar is more likely to become a coronal affricate than the voiced velar. For example, in English, Bantu and Slavic the voiceless velar has become a coronal affricate while the voiced velar has become a coronal fricative or palatal glide. Also, in all the examples given in table 1, velar palatalization creates a sound not previously found in the languages' segmental inventory.

Laboratory investigation of the role of perception in velar palatalization is also motivated by work in the phonetic literature. Velars before front vowels and palatoalveolars have an acoustic and perceptual affinity. A voiceless velar before a high front vowel is similar to a palatoalveolar fricative/affricate in terms of peak spectral frequency, and duration of aperiodic noise.

The burst of a velar stop before front vowels and the palatoalveolar fricative [ʃ] both have a major spectral peak between 2 and 4 kHz [Fischer-Jørgensen, 1954; Hughes and Halle, 1956; Halle et al., 1957; Strevens, 1960; Zue, 1976; Nartey, 1982;

Behrens and Blumstein, 1988; Keating and Lahiri, 1993]. The velar place of articulation has a longer release than bilabial and dental/alveolar [Lisker and Abramson, 1964; Klatt, 1975; Tekieli and Cullinan, 1979; Kewley-Port, 1983; Keating, 1984; Krull, 1991] as well as the longest frication interval [Fischer-Jørgensen, 1954; Fant, 1973]. Voiceless velar stops have VOT values that are almost as long as the frication interval of palatoalveolar fricatives [Tekieli and Cullinan, 1979]. In addition, the length of frication and aspiration is a function of the following vowel – there is longer frication and aspiration before a high vowel [Klatt, 1975; Tekieli and Cullinan, 1979; Ohala, 1983].

Perceptual studies have shown that the release burst of a velar stop is more heavily weighted as a perceptual cue to place of articulation than are the following formant transitions [Dorman et al., 1977; Smits, 1995]. Likewise, [ʃ] (and [s]) are perceived mostly on the basis of the noise portion and not the following formant transitions [Harris, 1958; Heinz and Stevens, 1961; Whalen, 1991]. Thus, it is possible that confusion of velar burst with a coronal burst would not be overridden by differing formant transitions. Indeed, velars before front vowels are often incorrectly heard as coronals in perceptual tests [Winitz et al., 1972; Repp and Lin, 1989].

Based on reviews of the typology of palatalization, we know that velars commonly palatalize before front vowels, and that high front vowels are preferred over mid or low front vowels [Chen, 1973; Neeld, 1973]. We also know that one of the most common outcomes of velar palatalization is a palatoalveolar affricate [Bhat, 1978; Hock, 1991]. As mentioned earlier, voiceless velars are more likely to palatalize than voiced velars [Bhat, 1978]. Given this, the present study will focus on finding a perceptual motivation for a common manifestation of velar palatalization: velars becoming palatoalveolar affricates before a front vowel.

Looking for a perceptual component in velar palatalization does not deny that articulation plays a role in this sound change. Indeed, the origin may have an articulatory base in fronted velars. In fact Keating and Lahiri [1993, p. 90] tentatively propose that velars before high front vowels will have a 'smaller-than-expected' front cavity due to the more anterior lateral contact of the following vowel. The smaller front cavity would produce a higher F2 resonance and a 'good chance of affrication'. Thus, the articulatory characteristics of fronted velars have acoustic effects that should be important in the sound change of velar palatalization.

It is proposed here that a perceptual reanalysis of the acoustic properties of fronted velars occasions the change from fronted velar to coronal. The proposal is not purely perceptual, the articulatory characteristics of velars before high front vowels play a crucial role in creating the conditions needed for a perceptual reanalysis. Articulation comes into play in another area as well. There is more coarticulation in faster or connected speech [Kohler, 1990], thus, by hypothesis, velars will be more fronted in faster speech than in citation speech. Therefore, it is further proposed that *velar palatalization is conditioned by a perceptual reanalysis of faster speech.*

This proposal makes several predictions that can be tested by acoustic and perceptual procedures. These specific predictions about the acoustic characteristics of velars and palatoalveolars will be investigated in Experiment I:

- *Prediction 1.* Velars before front vowels are more similar acoustically to palatoalveolar affricates than to velars before back vowels.
- *Prediction 2.* The acoustic similarity between velars and palatoalveolars is greater before high front vowels than before mid and low front vowels.

- *Prediction 3.* Voiceless velars are more similar acoustically to palatoalveolars than voiced velars are.
- *Prediction 4.* The acoustic similarity of fronted velars and palatoalveolars is greater in faster speech than in citation speech.

The hypothesis that velar palatalization arises from a perceptual reanalysis of fronted velars in faster speech also leads one to expect that fronted velars will be perceptually confusable with coronals. This prediction can be broken down into three parts. Experiment II tests the first prediction and Experiment III tests all three predictions.

- *Prediction 5.* Degrading the signal will produce velar/palatoalveolar confusions, with a vowel asymmetry whereby velars before high front vowels will more often be heard as palatoalveolars than velars before other vowels.
- *Prediction 6.* Faster speech productions of velars will more often be heard as palatoalveolars than citation productions.
- *Prediction 7.* There will be more [k]/[tE] confusions than [g]/[d/] confusions.

The hypothesis that velar palatalization is conditioned by a perceptual reanalysis of faster speech will receive support if these predictions are found to hold. If, on the other hand, the following outcomes are found, it would disconfirm the hypothesis.

- (1) Velars before front vowels are *not* more acoustically similar to palatoalveolars than to velars before back vowels.
- (2) Fronted velars in faster speech are *not* more acoustically similar to palatoalveolars than fronted velars in citation speech are.
- (3) Velars before high front vowels are *not* heard as palatoalveolars in the perception experiments at a greater rate than the velars before back and low vowels.

Experiment I

The experimental work reported in this section examines the acoustic similarity of velars and palatoalveolars. The spectral properties of the consonants and the second formant transitions of the vowels are investigated.

Method

Speakers

Seven native English speakers (4 female, 3 male) participated in this study. The speakers had no known hearing loss or speech impairment. All speakers were born and raised in Texas, were from 20 to 33 years old, and had at least some university education.

[1] The vowel broadly transcribed here as [u] is produced as a central vowel [ʉ]. High F2 values indicative of fronting were found for all speakers in this study. For example, a typical female speaker had a mean F2 value of 2056 Hz for citation speech and 2110 Hz for faster speech; and a typical male speaker had a mean F2 value of 1883 for citation speech and 1982 for faster speech. These are much higher than the values of 870 Hz for male and 950 Hz for female reported in Peterson and Barney [1952].

[2] The vowels [ɑ] and [ɔ] were merged by most of the speakers, as is common in Texas. The sequences [dʒʊ] and [tʃʊ] were not used since no English words begin with these combinations.

Stimuli

Forty-two word-initial consonant-vowel (CV) combinations of [k g tʃ dʒ] and [i ɪ eɪ ɛ æ ɑ ɔ oʊ ʊ u[1] ʌ] were collected from real English monosyllabic words.[2] The words ended in voiceless obstruents, except for the [tʃu] and [dʒɔ] sequences. Since no words end in voiceless obstruent for these sequences, the words *chew* [tʃu] and *josh* [dʒaʃ] were used. For example, the [ki] sequence was produced in the word *quiche.* A full list of the words used can be found in Appendix I.

Four tokens of each CV were recorded under two conditions: faster speech and citation speech. Two speech styles were used to determine whether or not the faster speech tokens of the velars would be more like palatoalveolars than the citation speech tokens. The faster speech condition was recorded first. The subjects did not know the purpose of the experiment. They were asked questions designed to elicit faster, reduced forms of the target words containing the CV sequences. The target words were presented with two different modifiers. The subject was asked the difference between a target word modified by *x* and a target word modified by *y.* Note that the stress is on the word modifying the target word, and not the target word itself. For example the subject was asked 'What's the difference between Snow Geese and Canada Geese' to elicit tokens of *geese* (containing the [gi] sequence). The subjects were instructed to repeat the question as a statement and then to answer the question. In this way, four tokens of the target word were obtained. This elicitation technique yielded relatively reduced forms. Appendix II contains a full list of the questions asked.

During the citation condition, the subjects were asked to read the same words in the frame 'Say ___ again'. Four tokens of each word were presented to the subjects on note cards that had been randomized by shuffling. They were instructed to speak clearly as if to someone across the room.

The speakers were recorded in a quiet, sound-attenuated room using a high-quality microphone with a 2-inch mouth-to-microphone distance on an audiocassette tape recorder.

Acoustic Analysis

The 2,352 tokens were digitized at 22.05 Hz (16 bit). The peak frequency of the release of the consonants and the effect of the consonant on the following vowel were analyzed. Two measurements were made for each token: the peak spectral frequency of the release burst and aspiration/frication of the consonants and the second formant (F2) vowel transitions. Peak spectral frequency was measured because bursts are known to be important in velar stop identification [Dorman et al., 1977; Smiths et al., 1996] and peak frequency of the burst is a good indicator of velar fronting [Fischer-Jørgensen, 1954; Halle et al., 1957; Fant, 1973; Zue, 1976]. The F2 transitions were measured so that velars and palatoalveolars could be compared using locus equations [Sussman et al., 1991].

The measures were made as follows: (1) *Peak Spectral Frequency:* An average spectrum of the burst/aspiration of the stops and of the frication portion of the affricates was made. The average spectrum was calculated with 512 FFT points and a 184-Hz filter. A +6 dB preemphasis of the higher frequencies was used. The frequency with the highest amplitude in the average spectrum was recorded. The peak frequency was determined over the whole range of the spectrum. The highest amplitudes were found in the 3- to 5-kHz range for velars before front vowels and palatoalveolars and in around 2 kHz for the velars before back vowels. (2) *Second Formant Transitions:* The frequency of F2 at the first discernible glottal pulse after the release burst and aspiration/frication (called F2C) was recorded. This measurement was made from a cursor reading taken from a spectrographic display. Spectrograms were generated with 256 FFT points, a 300-Hz filter, a frame advance of 0.1 ms, and +6 dB preemphasis of the higher frequencies. A time-locked waveform and expanded time display were used to help locate the first glottal pulse. F2C was measured at the first glottal pulse for stops and affricates as well as for voiced and voiceless tokens. This allowed a direct comparison of voiced F2 transitions across manner types. The frequency of F2 at the vowel midpoint (F2V) was measured from spectral displays. Following Sussman et al. [1991], if the formant resonance was steady state or monotonically changing in frequency, F2V was taken at the temporal midpoint. If, on the other hand, the F2 was U-shaped or the inverse, the measurement was taken at the minimum or maximum frequency respectively. LPC and FFT spectra were calculated with windows centered at the F2V measurement location. The LPC was calculated with 25 coefficients and a 20-ms frame length. The FFT spectra were calculated as described for measure (1). If the F2 in the LPC and FFT spectral peaks were aligned (i.e. the same), the measurement was taken of the F2 peak shared by the two types of spectra. If the F2 peaks were not aligned, the cursor was moved a few milliseconds forward and/or backward in the waveform window and new spectra were calculated. If the F2 peaks were still not

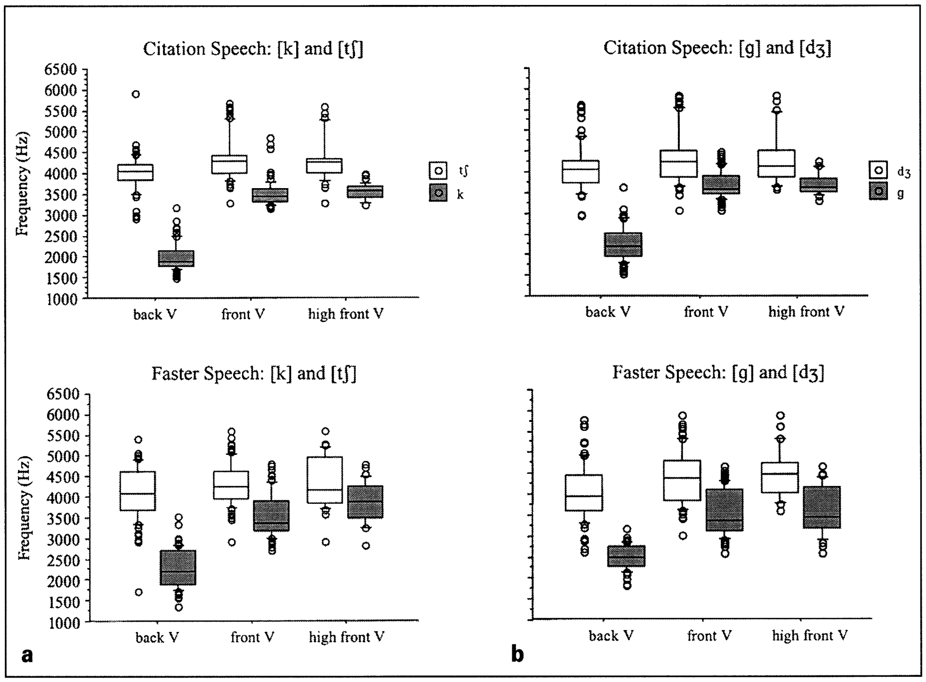


Fig. 1. Distribution of the peak spectral frequencies of [tʃ] and [k] (**a**) and [dʒ] and [g] (**b**) as spoken by the female speakers in the context of back vowels [u ʊ oʊ ɔ ʌ ɑ], front vowels [i ɪ e ɛ æ], and high front vowels [i ɪ] in two speech styles: citation and faster. The boxed area represents the range from the 25th to the 75th percentile. The line in the box represents the median and the lines above and below the box represent the 90th and the 10th percentile respectively. The remaining 20% of the distribution, or the outliers, are represented by the circles.

aligned the FFT and the LPC peaks were averaged. The averaging method was only applied in a handful of cases. The difference between the F2 peaks was never more than about 25 Hz and did not result from the LPC missing the formant entirely.

Results

All consonants were measured from release of closure to onset of vowel as defined by the beginning of the first regular glottal period in the waveform. The faster/citation condition produced a durational effect. The productions from each of the 7 speakers for all four consonants [k, tʃ, g, dʒ] showed the same trend: the faster speech tokens were shorter than the citation speech tokens. On average, the faster tokens were 29% shorter than the citation tokens (average citation duration was 80 ms and average faster duration was 57 ms).

Peak Spectral Frequency

No normalization was performed on the data to compensate for the expected effects of vocal tract size in the female and male subjects, so the male and female subjects will be treated as two separate groups. The data for each individual speaker show

Table 2. 2×2 ANOVAs on peak spectral frequencies of consonantal tokens from female speakers

Factors	Condition	F-value of interaction	
		citation speech	faster speech
Voice (voiceless vs. voice) × place (velar vs. palatoalveolar)	Front vowels	[$F(1,19) = 1.19$]	[$F(1,19) = 0.35$]
	High front vowels	[$F(1,7) = 0.72$]	[$F(1,7) = 5.47$]*
	Mid/low front vowels	[$F(1,19) = 0.46$]	[$F(1,19) = 1.22$]
Vowel height (high vs. mid/low) × place (velar vs. palatoalveolar)	Voiceless	[$F(1,18) = 0.35$]	[$F(1,18) = 5.90$]*
	Voiced	[$F(1,18) = 0.99$]	[$F(1,18) = 0.10$]

* $p<0.05$.

the same trends [Guion, 1996]. Therefore the data for the male speakers are pooled together and the data for the female speakers are pooled together.

The results for the female speakers will be presented first. Figure 1 shows the distribution of the peak spectral frequency for [tʃ] and [k] (a) and [dʒ] and g (b) in citation and faster speech. For [k] and [tʃ] (a), in both speech styles and voicing conditions, minimal overlap is observed between the peak spectral frequencies for velars and palatoalveolars before back vowels, but there is overlap before front vowels. In the environment before high front vowels the overlap is even greater. In the faster condition, overlap of the peak spectral frequencies for [tʃ] and [k] is greater before the high front vowels than before front vowels.

The distributions of the peak spectral frequencies for the voiced velar and palatoalveolar (b) show a similar pattern. There is no overlap in the distribution of the peak spectral frequencies for [g] and [dʒ] before back vowels, but thre is overlap before front vowels. Furthermore, ovrlap between [g] and [dʒ] before front vowels is greater in faster speech than in citation forms. Also note that velars before front vowels (both voicing types) have peak spectral frequencies that are more similar to the palatoalveolars than to the velars before back vowels, supporting prediction 1.

A repeated measures ANOVA with the factors speech style (citation or faster) × vowel context (front V vs. back V) × consonant [tʃ, dʒ, k, g] × repetition (4 levels) was computed to determine if peak spectral frequency differed significantly across the conditions. Main effects of vowel context [$F(1,19)=480.33$] and consonant type [$F(3,57)=164$] were significant. The effect of repetition was nonsignificant, indicating that the four productions of each type were similar. The two-way interactions of speech style × vowel context [$F(1,19)=5.87$] consonant type × vowel context [$F(3,57)=82.51$] as well as the three-way interaction of speech style × consonant type × vowel context [$F(3,57)=2.80$] were all significant, indicating that vowel context and speech style play a significant role in determining the peak spectral frequency of the consonants [tʃ dʒ k g].

A series of ANOVAs were conducted to explore these interactions. Since repetition was not significant, the ANOVAs described below were performed on data averaged across the four repetitions. F-values can be found in table 2. Repeated measures ANOVAs with the factors of Voice (voiced vs. voiceless) × Place (velar vs. palatoalveolar) were conducted on the front vowel data to test the prediction that voiceless velars are more like palatoalveolars than voiced velars are. The interactions were nonsignifi-

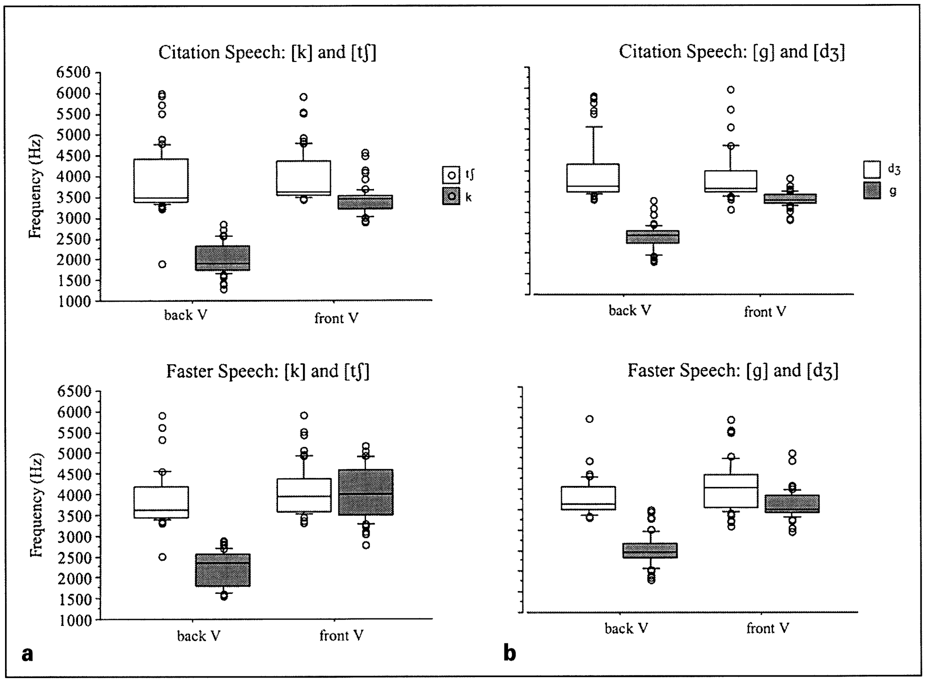


Fig. 2. Distribution of the peak spectral frequencies of [tʃ] and [k] (**a**) and [dʒ] and [g] (**b**) as spoken by the male speakers in the context of back vowels [u ʊ oʊ ɔ ʌ ɑ] and front vowels [i ɪ e ɛ æ] in two speech styles: citation and faster. The boxed area represents the range from the 25th to the 75th percentile. The line in the box represents the median and the lines above and below the box represent the 90th and the 10th percentile respectively. The remaining 20% of the distribution, or the outliers, are represented by the circles.

cant for both faster and citation speech. The voice×place interaction was further investigated by running ANOVAs on the *high* front vowel data. This time, the interaction was significant in the faster speech condition, with [k] being more like [tʃ], than [g] like [dʒ]. This finding indicates that, in faster speech before high front vowels, voiceless velars are more like palatoalveolars than voiced velars are, supporting prediction 3.

A second set of ANOVAs with the factors of vowel height (high vs. mid/low)×place (velar vs. palatoalveolar) was conducted on the *front* vowel data. These ANOVAs test the prediction that the acoustic similarity between velars and palatoalveolars is greater before high front vowels than before mid and low front vowels. The interactions were nonsignificant for the voiced condition in both citation and faster speech. In the voiceless condition, however, the interaction was significant in faster speech, with [k] being more like [tʃ] before high front vowels. These results indicate that [k] in faster speech is more like [tʃ] before high front vowels than before mid and low front vowels, supporting prediction 2.

As shown in figure 2, the peak spectral frequencies from the male speakers show the same trends seen in the female speakers. There is virtually no overlap in either speech style between [k] and [tʃ] or [g] and [dʒ] in the environment before back vowels. Before the front vowels, however, there is considerable overlap and that overlap is greater in the faster condition than in the citation condition. Again, like the female speakers, the velars before front vowels are more similar to the palatoalveolars than to the velars before back vowels. A repeated measure voice (voiced vs. voiceless) × place (velar vs. palatoalveolar) ANOVA performed on the front vowel data in faster speech had a significant interaction [$F(1,14)=8.20$], with [k] being more like [tʃ] than [g] like [dʒ]. The same ANOVA performed on the citation speech data returned a nonsignificant finding. These results indicate that voiceless velars are more like palatoalveolars than voiced velars are. In addition, that similarity is greater in faster speech.

Second Formant Transitions

The F2 transitions of the velars and palatoalveolars were analyzed using locus equations, which are generally considered reliable measures of place of articulation [but see Fowler, 1994, and Sussman and Shore, 1996]. Locus equations define a linear relation between F2 onset and midpoint for vowels following a consonant in a given place of articulation [Lindblom, 1963; Krull, 1988, 1989, 1991; Sussman et al., 1991, 1993, 1995]. Locus equations were derived by plotting F2 at the first glottal pulse and F2 at the vowel midpoint for a single place of articulation. A regression line was then fit to the scatter plot of all F2C-F2V coordinates across vowel contexts. It is the slope of the regression line that provides an index of coarticulation: the greater the slope, the greater the coarticulation. The locus equations for a given consonant have a greater slope (i.e., have flatter formant transitions) in faster speech than in citation speech, indicating more coarticulation and/or formant undershoot [Krull, 1989; Duez, 1989, 1990; Poch-Olivé et al., 1989].

Velars often show two distinct slopes and y-intercepts based on vocalic context: one group for the front vowels and one for the back vowels [see Sussman et al., 1991]. In other words, they do not show a single linear function across vowel types as other places of articulation do. In the current study, two regression equations (one for the front vowels and one for the back vowels) were computed for the velars. The points had a tighter fit to the regression lines when plotted separately by vowel context. Pooling the data for all speakers, the root mean square (RMS) residual of the linear prediction was significantly greater for the velars treated as a group than for the velars split by vowel context [$T(84)=-3.768$, $p<0.01$]. Splitting the palatoalveolars by vowel context, however, did not produce the same result. The difference in RMS residual for the palatoalveolar locus equations split by vowel type and the palatoalveolar locus equations for all vowels is not significant [$T(84)=-0.953$]. Therefore the locus equations presented here will split the velars by vowel context, but not the palatoalveolars.

Locus equations are sensitive to variations in speech rate and, in the case of velars, vowel effects. Therefore, they are an ideal measure to investigate velar fronting across speech styles and vowel contexts. The locus equations presented here were made with measurements of F2 at the first glottal pulse (F2C) and F2 at the vowel midpoint (F2V). Thus, when the locus equations of [k] were compared to [tʃ], for example, only the information found in the voiced portion of the vowel was being compared. Due to the difference in voice onset time (VOT), the locus equations for [k] and [g] are expected to be somewhat different. Since [k] has a longer VOT than [g], the tongue has a longer time to approximate the vowel target by the onset of voicing than it has for [g]. For this study the voiceless [k] and [tʃ], and the voiced [g] and [dʒ] will only be compared to each other.

Sussman and Shore [1996] describe a procedure for taking the F2C measurement for voiceless stops at roughly the same temporal point as the voiced stops. The F2 res-

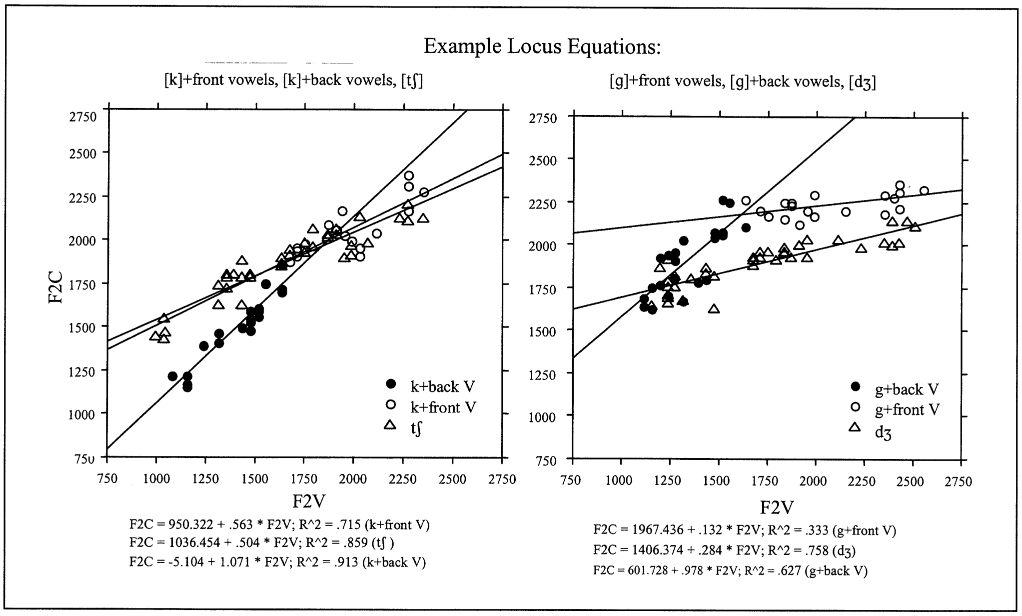


Fig. 3. Locus equations for [k] in a back vowel context, [k] in a front vowel context, and [tʃ] before all vowels (**a**), and for [g] in a back vowel context, [g] in a front vowel context, and [dʒ] before all vowels (**b**) for the citation forms of a single speaker.

onance is measured in the aspiration interval on a high-resolution spectrographic screen display. The earliest visible point of the F2 resonance (after the release burst) that continues the voiced resonance is measured. These data were also measured using the 'aspiration methodology'. Results of locus equations made with these measurements will be presented in the discussion section.

First consider the locus equations made from measurements at the first glottal pulse of the vowel. Figure 3 presents the regressions computed from locus equations for voiced and voiceless consonants for a single speaker. The locus equations for [k]+front vowels, [k]+back vowels and [tʃ] have been placed in the same scatterplot (a). F2C is plotted along the y-axis and F2V is plotted along the x-axis. Note that the regression lines for the [k] before front vowels and [tʃ] are similar, while the regression line for [k] before back vowels diverges. In (b) the locus equations for [g]+front vowels, [g]+back vowels and [dʒ] are presented. Note that the velar before front vowels and the palatoalveolar regression line are not as similar in the voiced case.

The slopes (x-axis) and y-intercepts (y-axis) for all 7 speakers are presented in figure 4. The locus equations for faster and citation speech are pooled in the graphs to provide a view of the overall distribution of the data. In (a), the distribution of [k] before front vowels and [tʃ] overlap almost completely, but the points for [k] before back vowels (filled circles) are distributed somewhat differently. In (b), the distribution for [g] and [dʒ] is shown. Note that distribution for [g] before back vowels is quite diffuse and overlaps with the distribution of [g] before front vowels. In this case, [dʒ] is split off from the other groups.

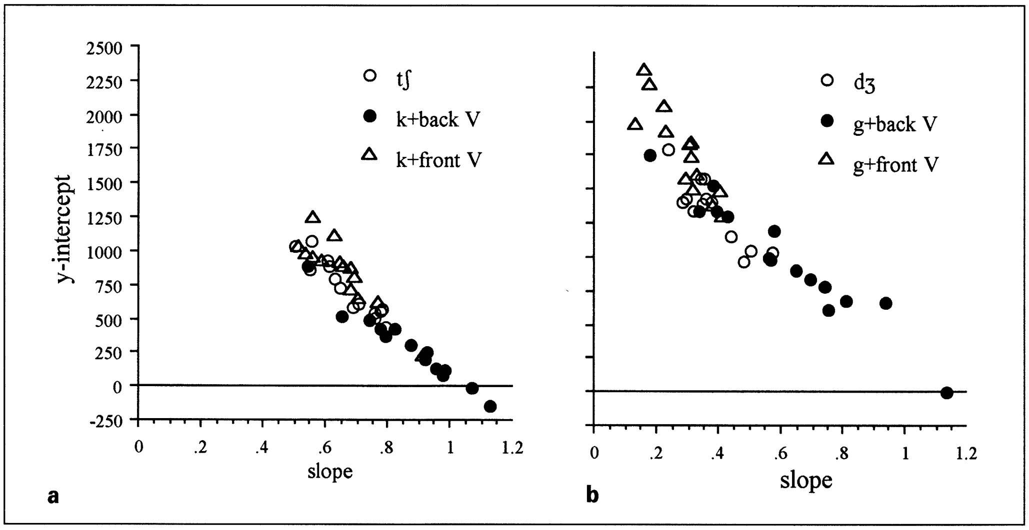


Fig. 4. Slope and y-intercept from the locus equations of all the 7 speakers. **a** y-Intercept by slope for the locus equations of [k] before front vowels, [k] before back vowels, and [tʃ] before all vowels. **b** y-Intercept by slope for [g] before front vowels, [g] before back vowels, and [dʒ] before all vowels. The data from the two speech styles has been pooled in the figures.

Table 3. Paired t tests for slope and y-intercepts of locus equations from all 7 speakers

Paired comparison	Citation speech		Faster speech	
	slope	y-intercept	slope	y-intercept
k + back V vs. tʃ	T(6) = –4.50*	T(6) = 6.14*	T(6) = –3.28*	T(6) = 4.8*
k + front V vs. tʃ	T(6) = –1.54	T(6) = 0.00	T(6) = 0.48	T(6) = 2.7
k + front V vs. k + back V	T(6) = –2.50	T(6) = 3.29*	T(6) = –2.98	T(6) = 4.27*
g + back V vs. dʒ	T(6) = –4.84*	T(6) = 4.90*	T(6) = –0.23	T(6) = 6.05*
g + front V vs. dʒ	T(6) = –3.04	T(6) = 4.44*	T(6) = –2.73	T(6) = 4.13*
g + front V vs. g + back V	T(6) = –4.82*	T(6) = 5.30*	T(6) = –1.74	T(6) = 2.13

*p<0.01.

The data were subjected to a repeated measures MANOVA with two dependent variables of slope and y-intercept (the two variables are known to be correlated [Sussman et al., 1991]). Sussman et al. [1991] showed that locus equations effectively act to normalize differences in vocal tract size: they found no statistical difference for locus equations of a given place of articulation with respect to sex of the speakers. Given this, the male and female speakers will not be treated as two separate groups. Two within subjects factors were designated: speech style and CV context. The speech style had two levels: faster and citation. Six levels of CV context were defined: [k]+back vowels, [k]+front vowels, [tʃ]+all vowels, [g]+back vowels, [g]+front vowels, and [dʒ]+all vowels.

The main effects of speech style [Wilks=0.11396, p=<0.01, approx. $F(2,5)=19.44$] and CV context [Wilks=0.07189, $p<0.01$, approx. $F(10,58)=15.83$] were significant. The interaction of the two factors is also significant [Wilks=0.21600, $p<0.01$, approx. $F(10,58)=6.86$]. These results indicate that there are significant differences in the slope and y-intercept between the 6 CV contexts ([k]+back vowels, [k]+front vowels, [tʃ]+all vowels, [g]+back vowels, [g]+front vowels, and [dʒ]+all vowels) and that speech style has a significant interaction on the slope and y-intercepts of the same CV type.

Post hoc tests were performed to determine which factors contributed significantly. Paired t tests were used to compare the slope and y-intercept for two locus equation types in a given speech style. The slopes and y-intercepts for velar before front vowels and velar before back vowels were compared to each other as well as to the palatoalveolar. This comparison was performed within a given speech style and voicing type. The results from the post hoc tests are in table 3. Note that the contextual variants of velars were significantly different from each other in slope or y-intercept, or both. Velars before back vowels and palatoalveolars were also significantly different. Most interesting, however, was the comparison of velars before front vowels to the palatoalveolars. In the case of the voiceless consonants, the fronted [k] did not differ significantly from the [tʃ] in slope or y-intercept in either speech style. The [g], however, was significantly different in y-intercept from [dʒ] in both speech styles.

Discussion

The acoustic data from Experiment I generally support the predictions derived from the hypothesis that velar palatalization results from perceptual reanalysis of faster speech. The predictions are repeated below for convenience:

- *Prediction 1.* Velars before front vowels are more similar acoustically to palatoalveolar affricates than to velars before back vowels.
- *Prediction 2.* The acoustic similarity between velars and palatoalveolars is greater before high front vowels than before mid and low front vowels.
- *Prediction 3.* Voiceless velars are more similar acoustically to palatoalveolars than voiced velars are.
- *Prediction 4.* The acoustic similarity of fronted velars and palatoalveolars is greater in faster speech than in citation speech.

The peak spectral frequency data support all four predictions. The peak spectral frequencies of velars (both voiced and voiceless) before front vowels were more similar to palatoalveolars than to velars before back vowels. For the males, the spectral peaks of [k] and [tʃ] before front vowels were more similar in faster speech than in citation speech. For the females, the spectral peaks of [k] and [tʃ] before high front vowels were more similar in faster speech than in citation speech. In addition, [k] was more like [tʃ] before high front vowels than before mid and low front vowels. Thus, we find a trend in which the spectral peaks of voiced and voiceless velars before front vowels are more like palatoalveolars than they are like velars before back vowels, and, in faster speech before front vowels (or high vowels in the case of the female speakers), voiceless velars are more like palatoalveolars than voiced velars are.

Results from the locus equation analysis reveal that when velar stops are compared to palatoalveolar affricates, [k] before front vowels and [tʃ] are similar, whereas [g] before front vowels and [dʒ] have more distinct distributions. The velars (both [k] and [g]) also show two distinct distributions as a function of vowel context. This supports prediction 1: velars before front vowels and palatoalveolars are more acoustically similar than velars before front vowels and velars before back vowels. It also supports prediction 3 since voiceless velars and palatoalveolars are more similar than the voiced velars and palatoalveolars. Prediction 4 was not supported because no difference was found between the two speech styles: [k] before front vowels and [tʃ] were not significantly different in either citation or faster speech.

One might suggest that the greater difference between voiced velars and palatoalveolars than between voiceless velars and palatoalveolars is due to the measurement of F2C at the first glottal pulse. As described in the method section, the decision to take F2C at the first glottal pulse for all four consonants was made in order to compare directly the information available the listener in the voiced vocalic portion of the syllable for both the velars and the palatoalveolars. Nonetheless, the measurement is made nearer to the onset of the transition in the voiced velar. This could yield greater variation across place in the voiced segments (i.e. the tongue would be in a more posterior position in the case of the voiced velar at the time of measurement). To answer this question, the F2C was remeasured for the voiceless velar at a location closer to the transition (i.e. in the aspiration as described by Sussman and Shore [1996]). This new measure will be called F2asp.

Locus equations were calculated for the voiceless velars before front vowels based on the F2asp measure. The slope and y-intercept from the locus equations were then compared to those for [tʃ] using t tests. Both the faster and citation conditions were compared, making a total of 4 tests. The only significant result obtained was for the comparison of slope in faster speech [$T(6) = -4.11$]. Thus, the [k] F2asp measure produced locus equations that were slightly more different from [tʃ] than the F2C measure (for which none of the 4 comparisons were significant). However, like the [k] F2C locus equations, the [k] F2asp locus equations are more similar to [tʃ] than the [g] locus equations are to [dʒ] (for which 2 of the 4 comparisons were significant).

Listeners can use the information in the burst to identify the following vowel [Winitz et al., 1972; Tekieli and Cullinan, 1979; Repp and Lin, 1989]. Given this, locus equations calculated using F2asp might more closely mirror the acoustic information used by the listener. The greater differentiation between [k] and [tʃ] found using the *F2asp* measure might reflect the listener's normal ability to hear the difference between fronted [k] and [tʃ]. The greater similarity between [k] and [tʃ] found using the *F2C* measure might reflect listener's tendency to hear [k] as [tʃ] when the burst/aspiration portion of [k] is masked as in Experiment III discussed below.

Experiment II

Given the acoustic similarity of fronted [k] and [tʃ] in terms of the noise portion and following formant transitions, one might expect to find a perceptual similarity as well. In Experiment II, tokens of [k] and [tʃ] were shortened to 30 ms and then presented to listeners for identification.

Table 4. Results from perception Experiment II for the two conditions

Heard	Spoken											
	All consonantal information						30 ms of the consonant					
	ki	kɑ	ku	tʃi	tʃɑ	tʃu	ki	kɑ	ku	tʃi	tʃɑ	tʃu
k	98	100	100	–	–	–	53	97	97	–	2	6
tʃ	2	–	–	100	100	100	47	3	3	100	98	94

Percentage [k], and [tʃ] responses to [ki], [kɑ], [ku], and [tʃi], [tʃɑ], [tʃu] stimuli.

Method

Speakers

Nineteen volunteers participated in this study. They were undergraduates, graduate students, and staff at the University of Texas and had no known hearing loss. All were native speakers of American English.

Stimuli

The stimuli consisted of tokens of [k] and [tʃ] spoken in the context of [i], [u], and [ɑ]. The tokens used were those from the faster speech of 2 speakers (1 female and 1 male) reported in Experiment I. There were 4 repetitions of each word type by each speaker. Thus, there were 8 different tokens for each of the 6 types (48 tokens in all).

The tokens were digitally edited in two ways. In the first condition (a), the voiced portion of the vowel and final consonant were deleted leaving only the burst and aspiration in the case of [k], and only the release and frication in the case of [tʃ]. In the second condition (b), all but the first 30 ms of the consonant was deleted. (All of the consonants were more than 30 ms long, so no voiced vowel portion remained.)

Procedure

The stimuli were presented to the 19 listeners at a comfortable level over headphones. The stimuli including the whole consonant, condition (a), were presented first. The stimuli were randomized and recorded onto audiotape in blocks of 10 with an interstimulus interval of 2 s. There were 8 s between blocks of 10 stimuli. The blocks were presented in two counterbalanced groups. Then the gated stimuli, condition (b), were presented, also in two counterbalanced groups. The subjects were given an answer sheet and told to circle which consonant of the two they heard in a forced choice. The choices k and ch were printed for each response on the answer sheet. There were a total of 1,824 responses, 912 responses per condition.

Results and Discussion

Table 4 (left half) shows the results of the first condition, in which all of the aperiodic noise was retained. The numbers given are percentages. The diagonal line running from top left to bottom right shows the percentage of correct identifications. Mistakes were made only for the [k] before the high front vowel [i]. The results for the second condition (in which only the first 30 ms of the tokens were played) are given on the right half of table 4. [k] before [ɑ] and [u] and [tʃ] before [i], [ɑ] and [u] were all identified correctly in more than 90% of the instances. The [k] before [i] on the other hand was identified correctly only 53% of the time. This is slightly above chance. The tokens beginning with [k] and the high front vowel [i] were highly confusable with the palatoalveolar affricate. The subjects appear to have been guessing between a response of [k] and [tʃ] for these tokens.

The results from Experiment II supported the prediction that [k] before front vowels would be confusable with [tʃ]. The confusion was uni-directional: [k] was heard as [tʃ], but [tʃ] was *not* heard as [k]. This directionality paralles 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. Because there were only two choices in the paradigm, we do not know if the 47% [tʃ] responses to [ki] was due to random identification of an unknown segment, or if [ki] does indeed sounded like [tʃ].

Experiment III

The results from Experiment II suggested that [k]s before high front vowels may be confusable with [tʃ]. However, a methodology with more than two response alternatives was needed to determine if the results were due to random guessing or to genuine [k]/[tʃ] confusion. A larger experiment was also needed to investigate the role of voicing and speech style (citation vs. faster) on the consonant identification to test predictions 6 and 7.

Experiment III investigates the [k]/[tʃ] confusion further in an experiment with four possible responses. Speech style and voicing have also been factored into the design. This time the signal was degraded by using masking noise. This design feature was prompted by several observations. First, it was hoped that the noise condition would better mirror real-world listening conditions. (Noisy conditions are quite common in everyday conversation whereas sounds truncated to the first 30 ms are rare.) Of course, one might argue that the noise needed to induce many perceptual errors is greater than that which ocurs in real-world listening conditions. Second, the noise masking condition, unlike the gating condition, allows investigation into durational cues such as VOT.

Method

Subjects

Twenty-four volunteer subjects participated in this study. All were undergraduates in linguistics courses at the University of Texas at Austin and had no known hearing loss. All were native speakers of American English.

Stimuli

The stimuli for Experiment III were taken from faster and citation productions of words beginning with the consonants [k tʃ ɡ dʒ] followed by the vowels [i ɑ u]. The productions of all 7 speakers from Experiment I were used. Thus there were 24 types spoken four times each by 7 speakers (total = 672).

The stimuli were digitally edited to remove all but the consonant and 100 ms of the vowel. 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 waveform. The remainder of the word was truncated at a zero crossing in the waveform 100 ms inside the vowel. After editing, the stimuli were normalized for intensity. The RMS of the voltage was taken from a 50-ms window in the vowel. The left edge of the window was at least three glottal pulses inside the wovel. Then the whole demisyllable was multiplied by a factor to make the measured RMS equal to 2 V. 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. For the stimuli in the noise condition, white masking noise was added. The signal to noise ratio (S/N) was set at +2 dB. This level was determined by a series of pilot experiments that investigated S/Ns from +10 dB to –2dB in 2-dB increments. An S/N of +2 dB yielded many errors, but not random guessing.

Procedure

The stimuli were randomized (the stimuli with and without noise were kept in separate groups). The randomized demisyllables were recorded on audiotape in blocks of 10 with an interstimulus interval of 2 s. There were 8 s between blocks of 10 stimuli. The noise stimuli were also divided into three sets (A, B and C) to be presented to the subjects on a rotational basis.

The stimuli were presented to the 24 subjects. One hundred and twenty warm-up stimuli were presented first in which the S/N was decreased for each of three sets of 40 stimuli from +10 dB S/N to +6 dB S/N and finally to +2 dB S/N. Then subjects heard the three sets of noise stimuli. (Set A had 230 stimuli, Set B had 230 stimuli, and Set 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 told to circle one of four possible answers (k ch j or g) in a forced choice.

The responses to the no-noise stimuli were scored first and used to evaluate the subject's performance on the task. To be retained, a subject 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. The 4 (of 24) subjects who scored below 90% were dropped.

Twenty 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.

Results

Table 5 presents the overall results from the noise masked condition. The response consonant is listed in the far left column. The diagonal gives the percentage of correct identifications. Each consonant vowel sequence received more than 1,000 responses. The average correct response for the noise-masked stimuli was 69% correct, which is far above chance (25%). Table 6 presents the results for the no-noise stimuli. The percent correct rate is much higher than in the noise condition, the average being 96% correct. 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 unusually short VOT.

In the noise condition, [k] was heard as [tʃ] more often than any other consonant 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 [kɑ] 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 (25%) was for [dʒ] to be heard as [tʃ]. The third most common confusion (around 15%) was [g] heard as [dʒ]. More [g]/[dʒ] confusions occurred before the vowel [i].

Table 5. Results from perception Experiment III: noise-masked 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

Percentage [k], [tʃ], [g], and [dʒ] responses to [ki], [kɑ], [ku]; [tʃi], [tʃɑ], [tʃu]; [gi], [gɑ], [gu], and [dʒi], [dʒɑ], [dʒu] stimuli.

Table 6. Results from perception Experiment III: no-noise stimuli

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

Percentage [k], [tʃ], [g], and [dʒ] responses to [ki], [kɑ], [ku]; [tʃi], [tʃɑ], [tʃu]; [gi], [gɑ], [gu], and [dʒi], [dʒɑ], [dʒu] stimuli.

The results of Experiment III clearly support prediction 7: there were more [k]/[tʃ] than [g]/[dʒ] confusions. About 15% of the [g] tokens were heard as [dʒ], whereas around 26% of the [k] tokens were heard as [tʃ].

Prediction 5, that velars would be heard as palatoalveolars most often 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 before [i]. Perhaps these results can be attributed to the fronted production of [u] in the speech of the subjects (see footnote 1). These results differ from Experiment II, in which there was a clear split between [i] and [ɑ u] in terms of palatoalveolar response. Here, the high vowels function as a class in opposition to the low vowel. The differing results may reflect the effects of gating versus noise masking. In the truncated condition, durational effects were lost, whereas in the noise-masking condition, durational cues are kept. In both experiments, frequency information was available. However, the similarity of sibilant frication and the white noise used in the noise condition would seem to predispose an affricate response. Perhaps the higher average peak spectral frequency of [ku] as compared to [kɑ] coupled with a predisposition for an affricate response precipitated the greater [tʃ] response to [ku] stimuli.

Since the vowels [i] and [u] are intrinsically less intense than [a] [Lehiste and Peterson, 1959], it is possible that the normalization procedure (which was based on vowel intensity) produced [ki] and [ku] tokens with relatively more intense aperiodic noise than the [ka] tokens. The greater amplitude of the [k] before the [i] and [u] could lead to more [tʃ] responses. In order to determine if, indeed, the normalization procedure introduced a confounding variable, the post-normalized intensities of the aperiodic portions of [k] in all three vowel contexts were measured. RMS amplitude envelopes were calculated for all [k] stimuli using a 20-ms window. The peak intensity of the nonperiodic portion of the [k] was measured. The amplitude measure was then correlated with the number of [tʃ] responses. This correlation was nonsignificant ($r=-0.006$), indicating that the amplitude of the [k] stimuli did not introduce a response bias for [tʃ].

Prediction 6, 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 presented below indicate that speech style (citation vs. faster) 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, perhaps because of a ceiling effect. 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 a higher signal-to-noise ratio.

The number of palatoalveolar responses ([tʃ] and [dʒ] combined) to velar stimuli were subjected to a speech style (2)×voicing (2)×vowel context (3) ANOVA with repeated measures on all three factors. All main effects, as well the three-way interaction, were significant: speech style [$F(1,19)=8.68$], voicing [$F(1,19)=28.84$], vowel context [$F(2,38)=32.50$], and speech style×voicing×vowel context [$F(2,38)=6.34$]. This indicates that the speech style, voicing of the velar, and vowel context all have an effect on the percentage of palatoalveolar responses to velar stimuli.

In the following sections, the results of the noise perception test for the velar stimulus consonants are presented separately. Since the ANOVA examining the responses to the velars indicated that voicing had a significant effect, further tests were done for [k] and [g] separately, using the same basic design as the overall ANOVA. The data were split into two groups ([k] and [g]) and subjected to an ANOVA with the factors of speech style and vowel context.

Table 7. Results for noise-masked [k] stimuli

Heard	Spoken					
	citation speech			faster speech		
	ki	kɑ	ku	ki	kɑ	ku
k	45	89	50	42	79	43
tʃ	37	10	37	33	16	25
g	7	–	4	12	4	19
dʒ	11	–	9	13	1	13

Percentage [k], [tʃ], [g], and [dʒ] responses to [ki], [kɑ], and [ku] stimuli in faster and citation speech.

[k] *Stimuli*

The responses to [k] are listed in table 7. The responses to citation and faster speech stimuli have been separated. The overall proportion of [k] to palatoalveolar responses changed little in the two speech styles (it is 2.3 in citation speech and 2.2 in faster speech). The main difference was that the percentage of voiced [g] responses increased from 4% in citation speech to 12% in faster speech. There are more palatoalveolar responses to [ki] and [ku] thaan to [kɑ] in both speech styles.

A two-way repeated measures ANOVA examining the percentage of palatoalveolar responses to the [k] stimuli revealed no significant effect for speech style [$F(1,19)=0.62$]. The effect of vowel context was, however, significant [$F(2,38)=39.77$].

It is possible to investigate the relationship between the patterning of the responses and the acoustic attributes of the CV sequences since the tokens used in the perception experiment are the same as those used in the acoustic study. The VOT of the [k] tokens as well as the peak spectral frequency of the burst and aspiration of [k] were examined. The mean VOT of [k] was found to be 101 ms in citation speech and 64 ms in faster speech. The duration of [tʃ] (from release to vowel onset) overlaps with the VOT of [k]. The mean duration of [tʃ] in citation speech is 125 and in faster speech is 91 ms. Figure 5a illustrates the average VOT of the [k] stimuli used in Experiment III. Note that the citation stimuli have longer VOTs across the board and fall within the range of the [tʃ] means.

The VOT of [k] is clearly not the only factor contributing to the number of palatoalveolar responses. If it were, one would expect citation forms of [kɑ] to have the largest average palatoalveolar response (since they have the longest VOTs). But this was not 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, one would expect there to be more [tʃ] responses in the VOT range that

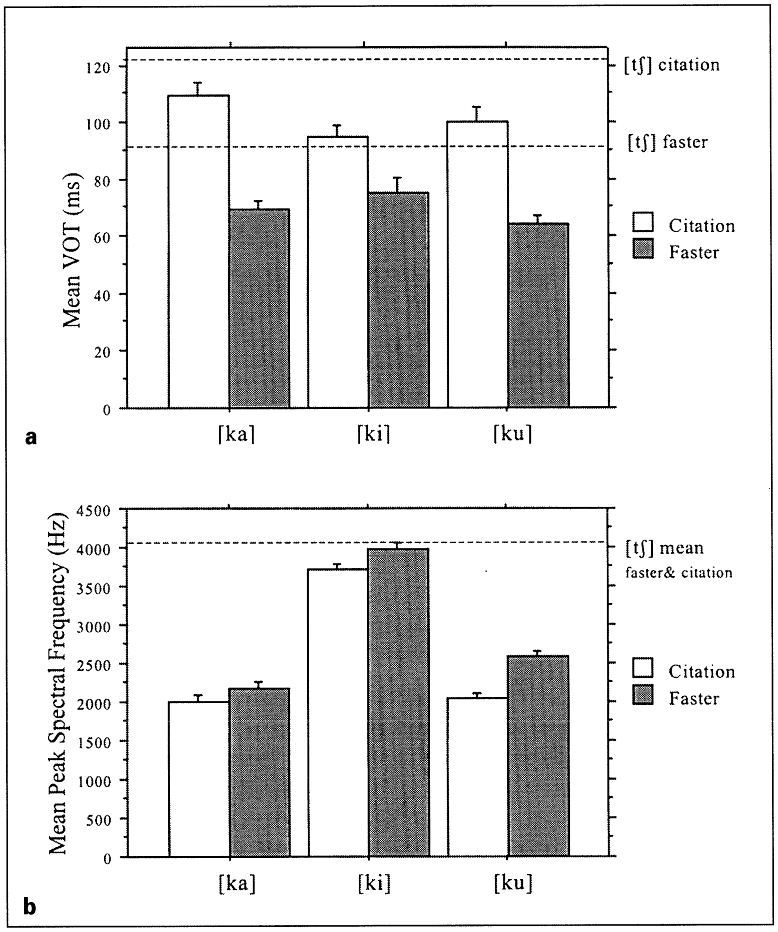


Fig. 5. **a** Mean VOT for [k] stimuli by speech style and vowel context. Dashed lines indicate the mean length of [tʃ] (before all vowel types) in citation and faster speech. **b** Mean peak spectral frequency of the burst and aspiration in [k] stimuli by speech style and a vowel context. Dashed line indicates mean peak spectral frequency of the frication in [tʃ] (before all vowel types) in citation and faster speech for both male and female speakers.

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.

Figure 6 illustrates the average number of responses to [k] (y-axis) as a function of VOT (x-axis). In the case of the [ki] stimuli in 6a, there is a clear trend of increased [tʃ] responses as VOT increases. The longest VOTs have an average [tʃ] response of almost 60%. 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 [$F(1,54) = 16.157$, $R^2 = 0.23$]. Also, [g] re-

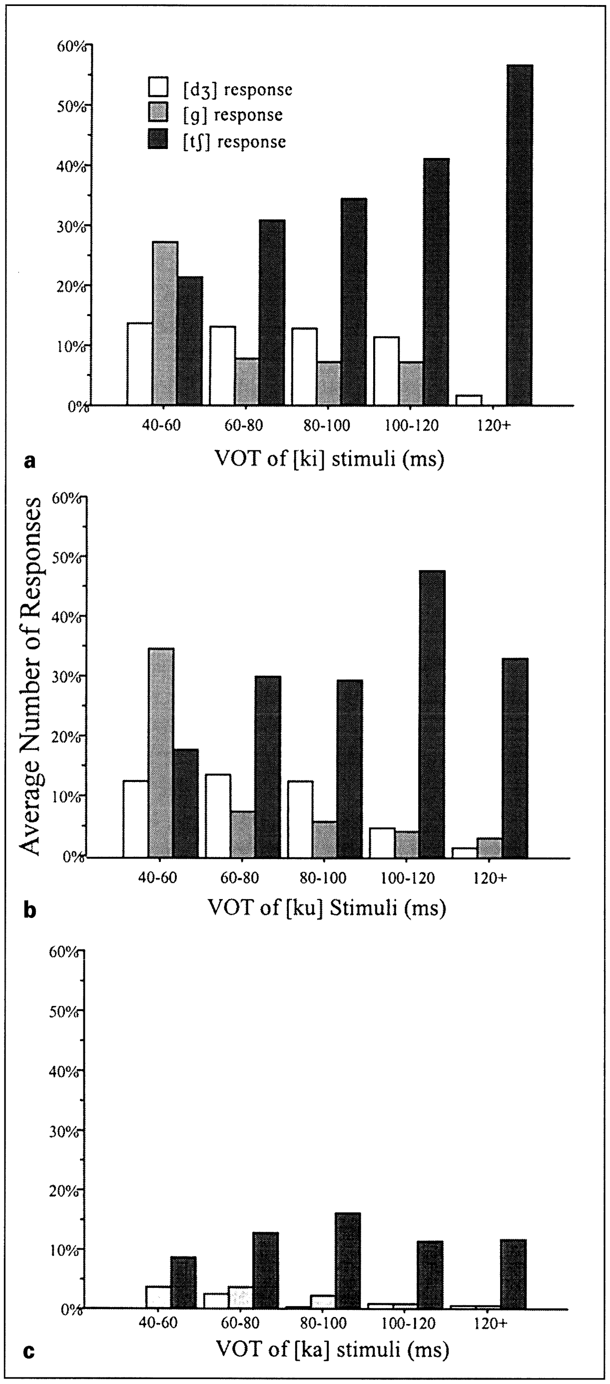


Fig. 6. The average number of [dʒ], [g], and [tʃ] responses to [ki], [ku], and [kɑ] stimuli. The horizontal scale groups the [k] stimuli in terms of a 20 ms. VOT range. The columns each represent a type of response. The white columns represent the average number of [dʒ] responses; the lightly shaded columns represent the average number of [g] responses; and the darkly shaded columns represent the average number of [tʃ] responses. Average number of responses is represented in terms of percent (out of a possible 20).

sponses decrease with increase of VOT. A linear regression analysis which has duration of [ki] VOT as an independent variable and average number of [g] responses as the dependent variable returns a significant finding [$F(1,166)=26.632$, $R^2=0.14$].

The responses to [ku] stimuli in 6b show the same basic trend as the responses to [ki]. Note that there are more [tʃ] responses when [ku]'s VOT is over 60 ms. 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 35% [g] responses. This number drops dramatically in the longer VOT ranges.

Finally, the [kɑ] stimuli in 6c have a much smaller overall incorrect response rate. The most common misidentification is [tʃ], but the average number of [tʃ] responses never goes above 20%. The trend towards more [g] responses in the shorter VOT ranges is weakly echoed in the [kɑ] data.

In summary, VOT had a different effect on the number of [tʃ] responses in the three vowel contexts. The [ki] stimuli showed a clear trend toward increasing the number of [tʃ] responses as the VOT increased. The [ku] stimuli showed a weaker trend, but the [kɑ] stimuli did not. 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 [tʃ] range for the VOT to have an effect on the number of [tʃ] responses, or perhaps some other factor not investigated here plays a role.

If the peak frequency of [k] release bursts influenced the number of [tʃ] responses, one would expect more [tʃ] responses when the peak spectral frequency of [k] was similar to that of [tʃ]. The mean frequency of [tʃ] (for both male and female) was 4067 Hz in citation speech and 4077 Hz in faster speech. Figure 5b illustrates the average peak spectral frequency of the [k] stimuli used in Experiment III. The faster stimuli have higher peak spectral frequencies across the board than the citation stimuli. 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 the result of the relatively fronted production of [u] in Texas English. The average peak spectral frequency of the faster [ki] stimuli comes closest to the [tʃ] mean.

Based on peak spectral frequency as shown in figure 5b, we would expect [ki] stimuli to be the most confusable with [tʃ]. The [ki] stimuli do receive the most palatoalveolar responses, but [ku] comes in a close second as can be seen in table 7. We do find, however, that the most palatoalveolar responses are found in roughly the range of the average [tʃ] peak spectral frequency (around 4000 Hz). Consider figure 7 and note that the greatest average [tʃ] response is to be found from 3500 to 4000 Hz. [tʃ] responses taper off at higher and lower frequencies. Here we have evidence that the acoustic similarities between [k] and [tʃ] which we found in Experiment I do play a role in perception. Those tokens of [k] that have a peak spectral frequency within the range of [tʃ] are more often confused with [tʃ].

As for the interplay between VOT and peak spectral frequency, we saw that the number of [tʃ] responses to [k] stimuli increased as VOT increased. This trend was the clearest in the [i] vowel context. Perhaps the peak spectral frequency of [k] needed to reach some critical level for the VOT effect on [tʃ] responses to be readily noticed. Figure 8 provides a graphical representation of the interaction of VOT and peak spectral frequency. First, note that as the peak spectral frequency of [k] increased the number of [tʃ] responses increased. Second, the number of [tʃ] responses increased as the

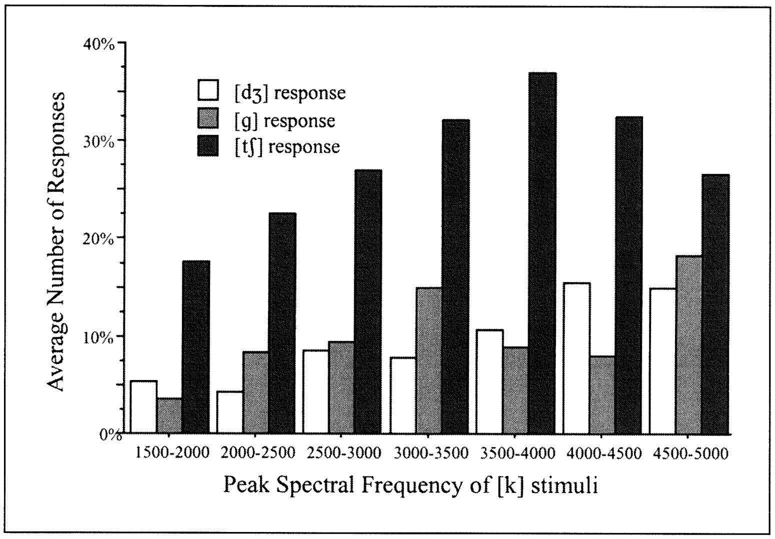


Fig. 7. Average number of responses (out of a possible 20) to [k] stimuli. The horizontal scale groups the [k] stimuli by a 500-Hz interval. The columns each represent a type of response. The white columns represent the average number of [dʒ] responses; the lightly shaded columns represent the average number of [g] responses; and the darkly shaded columns represent the average number of [tʃ] responses.

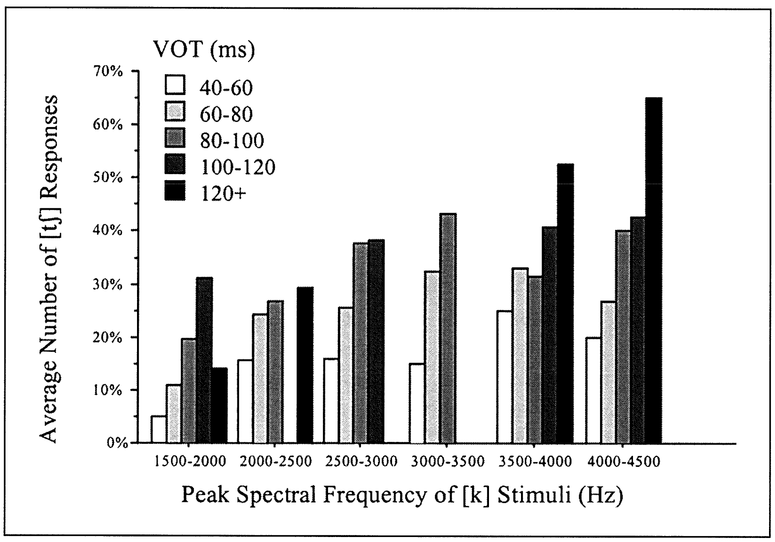


Fig. 8. Mean number of [tʃ] responses to [k] stimuli by peak spectral frequency and VOT of [k]. The x-axis lists the peak spectral frequency of the [k] stimuli. The length of the bars represents the average percent of [tʃ] responses (out of a possible 20) as defined by the numbers on the y-axis. For each frequency range, the responses have been split by VOT. The VOT is indicated by the legend in the upper left corner.

Table 8. Results for noise-masked [g] stimuli

Heard	Spoken					
	citation speech			faster speech		
	gi	gɑ	gu	gi	gɑ	gu
k	5	–	3	3	7	6
tʃ	6	–	3	2	2	3
g	64	91	77	79	82	76
dʒ	25	8	17	16	9	15

Percentage [k], [tʃ], [g], and [dʒ] responses to [gi], [gɑ], and [gu] stimuli in faster and citation speech.

VOT increased within each frequency range. Third, as the peak spectral frequency of [k] increases, the VOT plays a larger role in determining the number of [tʃ] responses. The [k] stimuli with the highest peak spectral frequency and the longest VOT received the most [tʃ] responses (almost 70%).

[g] *Stimuli*

The responses to noise-masked g stimuli are listed in table 8. Most of the errors lie in hearing [g] as [dʒ]. Also, [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]. A repeated measures ANOVA examining the percentage of palatoalveolar responses to the [g] stimuli revealed a signficant effect for speech style (citation vs. faster [$F(1,19)=17.52$] and for vowel context [$F(2,38)=9.20$], as well as the interaction of these two factors [$F(2,38)=13.82$]. These results indicate that speech style and vowel context combine to influence the number of palatoalveolar responses to [g] stimuli.

Analysis of the acoustic attributes of the [g] stimuli revealed that length of VOT and peak spectral frequency have an effect on the number of [dʒ] responses. The average VOT of the [g] stimuli is much shorter than the duration of frication in the palatoalveolars. The average duration of [dʒ] is 53 ms in faster speech and 71 ms in citation speech. As can be seen in figure 9a, the [gi] tokens in citation speech have the longest VOTs at around 33 ms. The stimuli with the longest VOT ([gi] citation) are the stimuli with the highest palatoalveolar response.

As seen in figure 10, [g] stimuli with VOT over 30 ms had an average [dʒ] response rate of over 20%. Those stimuli with a VOT of less than 30 ms had an average [dʒ] response rate of 15% or less. The trend of greater [dʒ] responses for longer VOT durations was found to be significant in a regression analysis, although only a small amount of the variance is accounted for [$F(1,166)=13.211$, $R^2=0.074$].

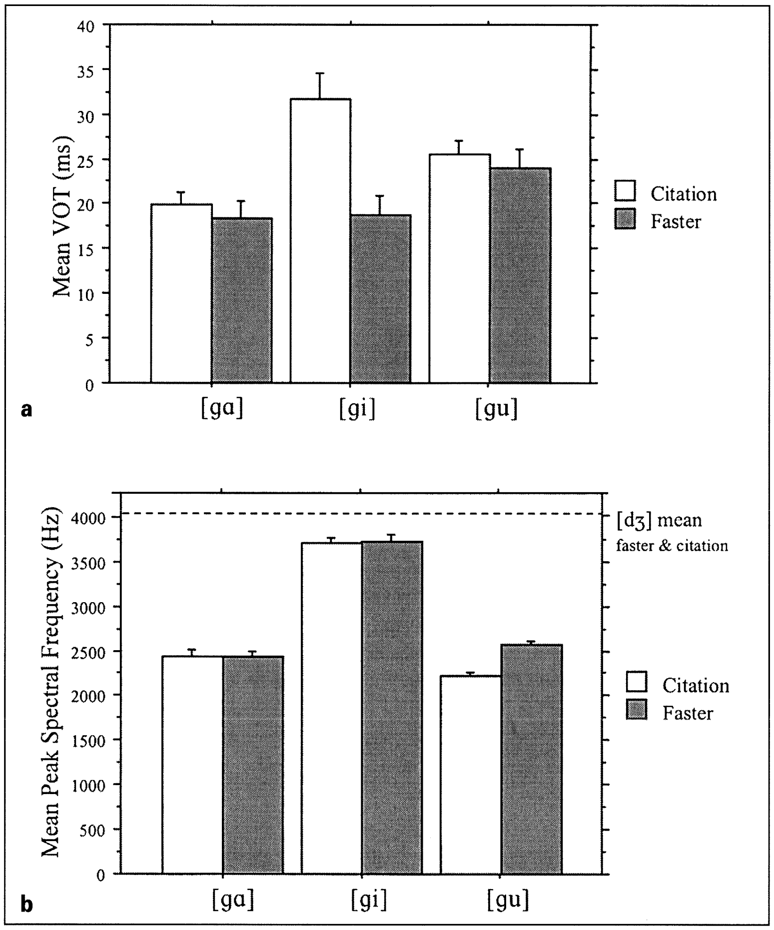


Fig. 9. **a** Mean VOT for [g] stimuli by speech style and vowel context. **b** Mean peak spectral frequency of the burst and aspiration in [g] stimuli by speech style and vowel context. Dashed line indicates means peak spectral frequency of the frication in [dʒ] (before all vowel types) in citation and faster speech for both male and female speakers.

The average peak spectral frequency for the [g] stimuli in general falls below the mean for the peak spectral frequency of [dʒ]. The mean peak frequency for [dʒ] was 4078 Hz for citation speech and 4100 Hz for faster speech. Figure 9b shows that the mean peak spectral frequency of [gi] stimuli came the closest to the mean for [dʒ]. Again, this is reflected in the number of palatoalveolar responses by vowel context: the [gi] stimuli were most often identified as palatoalveolars.

The peak spectral frequency of the [g] stimuli also has an effect on the number of [dʒ] responses. As figure 11 indicates, the higher the peak spectral frequency of [g] the greater the average [dʒ] response is. This relationship is significant [$F(1,157)=17.72$, $R^2=0.102$].

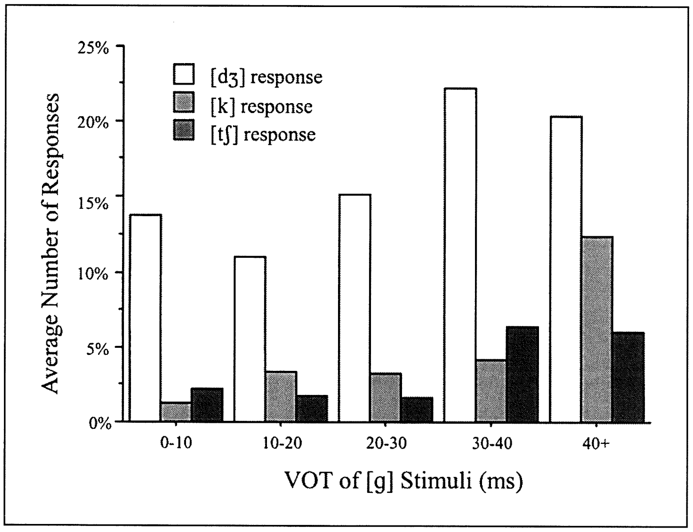


Fig. 10. The average number of [dʒ], [k], and [tʃ] responses to [g] stimuli. The horizontal scale groups the [g] stimuli in terms of a 10 ms VOT range. The columns each represent a type of response. The white columns represent the average number of [dʒ] responses; the lightly shaded columns represent the average number of [k] responses; and the darkly shaded columns represent the average number of [tʃ] responses. Average number of responses is represented in terms of percent (out of a possible 20).

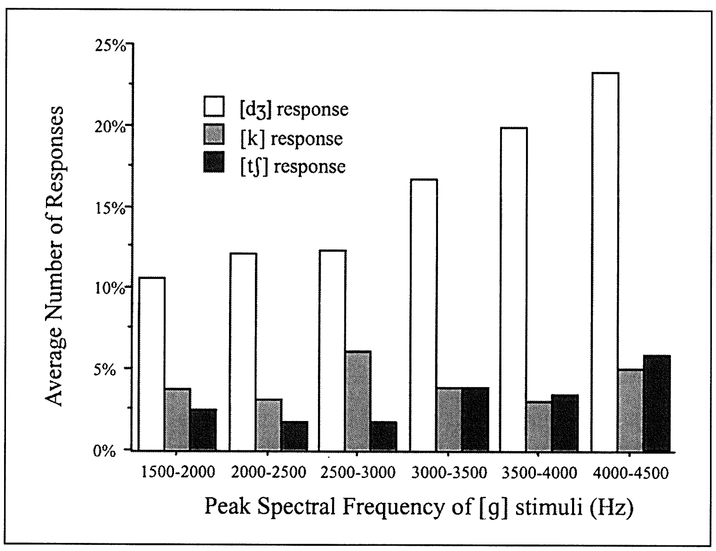


Fig. 11. Average number of responses (out of a possible 20) to [g] stimuli. The horizontal scale groups the [k] stimuli by a 500-Hz interval. The columns each represent a type of response. The white columns represent the average number of [dʒ] responses; the lightly shaded columns represent the average number of [k] responses; and the darkly shaded columns represent the average number of [tʃ] responses.

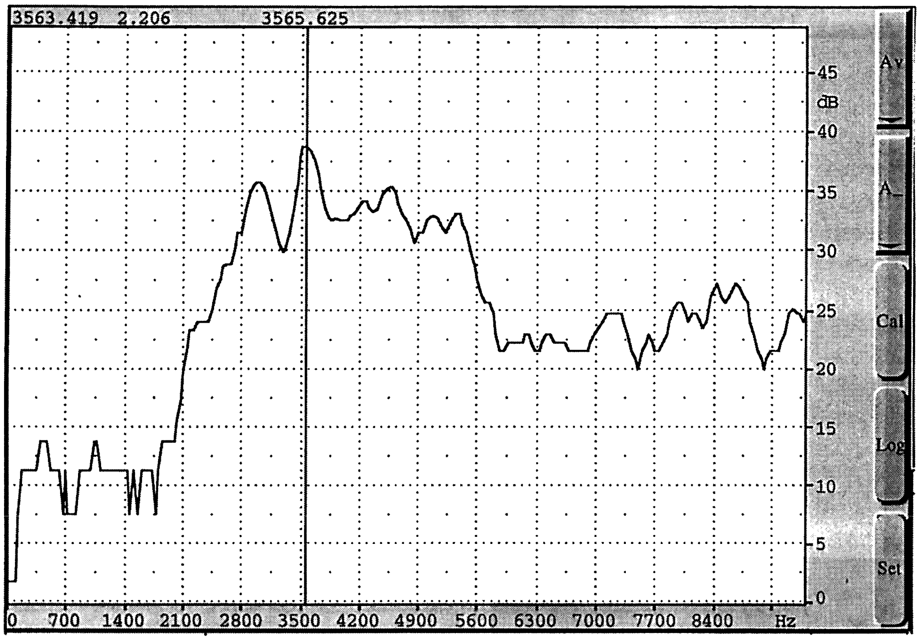
In summary, Experiments II and III supported prediction 5, that degrading the signal will produce velar/palatoalveolar confusion and that velars before high front vowels would most often be heard as palatoalveolars. In Experiment II, [ki] sequences were heard as [tʃ] about half the time while the [kɑ] and [ku] sequences were heard correctly over 90% of the time. In Experiment III, 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 lent itself to an affricate identification. The results also agree with prediction 7, that there would be more [k]/[tʃ] confusion than [g]/[dʒ] confusions.

Prediction 6, that faster speech tokens of [k] would be heard as [tʃ] more often than citation speech tokens of [k], was not supported, perhaps because the noise had a greater effect on the responses than the speech style. The speech signal was highly degraded by the masking noise. The gross effect of the noise seems to have eclipsed any faster/citation effect.

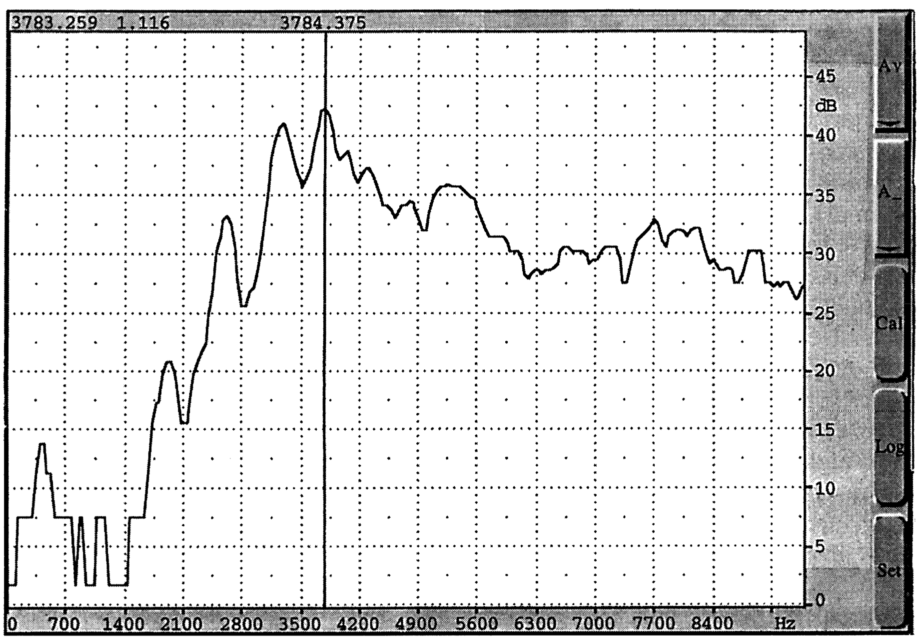
Summary and Conclusion

The experiments presented here investigated the proposal that velar palatalization is perceptually conditioned [Ohala, 1989, 1992]. More specifically, the hypothesis that *velar palatalization arises from a perceptually conditioned reanalysis of faster speech* was tested. The results of the experiments generally support the hypothesis that velar palatalization is perceptually conditioned. However, the hypothesis that the reanalysis is based on faster speech is only weakly supported. Six of the seven predictions laid out in the introduction received at least partial support and the three possible disconfirming scenarios were not found. In Experiment I, velar stops before front vowels and palatoalveolar affricates were shown to be acoustically similar in terms of peak spectral frequency and second formant transitions. In addition, the voiceless velars and palatoalveolars were more similar than the voiced consonants. This finding agrees with the most common target of the sound change, the voiceless [k]. Faster speech tokens of fronted velars were also shown to be more acoustically similar to palatoalveolars than citation speech tokens in terms of peak spectral frequency (but not in terms of their locus equations or durations).

In perception studies, velar stops before high front vowels were shown to be highly confusable with palatoalveolar affricates when the signal was degraded by noise masking or gating. The voiceless velars were misheard more often than the voiced velars. The results of the perception experiments complement the findings of the acoustic investigation. The consonants [k] and [tʃ] were found to be similar acoustically, especially before front vowels. In the perception experiments, [k] was often heard as [tʃ] when it was before a high front vowel. In Experiment II, there was a clear vowel effect whereby [k] before [i] was heard as [tʃ] most often. In Experiment III, [k] was often heard as [tʃ] before both [i] and [u]. It was proposed that noise masking and fronted production of [u] combined to produce this effect. Future perception experiments using other back vowels such as [oʊ ʊ ɔ] or nonfronted [u] could investi-



a



b

Fig. 12. The average long-term spectra for the burst and aspiration of [ki] (**a**) and the frication portion of [tʃi] (**b**) spoken by a single speaker in faster speech.

gate the effect of the fronted [u] on [k] perception in noise. If the other back vowels patterned more like [ɑ], then the large number of [tʃ] identifications for [ku] could be attributed to the fronted production of [u]. If on the other hand, the other back vowels patterned with the fronted [u], the large number of affricate responses could be due to the high noise level. In that case, further experimentation changing the signal to noise ratio and/or type of noise (e.g. using 'pink' noise) would be in order.

The predicted difference in [k]/[tʃ] perceptual confusion as a function of speech style was not found. There was no overall effect of speech style. Perhaps the signal to noise ratio was low enough (i.e. the noise was so loud) that it obscured the expected speech style effect. Further investigation with a higher signal to noise ratio could test this proposal. Since the peak spectral frequencies of [k] before front vowels were found to be higher in faster speech and spectral frequency is related to number of palatoalveolar responses (see fig. 7), further research into this area might find an effect of speech style. However, from the results of the ANOVA on the responses to [k] stimuli in Experiment III, we can only conclude that speech style had no effect on the number of palatoalveolar responses to velar stimuli, disconfirming prediction 6.

It is interesting to note that [tʃ] was rarely heard as [k]. The asymmetry found in the perception experiments is consistent with the sound change typology: a [k] becoming [tʃ] is quite common, whereas a [tʃ] becoming [k] is quite rare. Ohala [1985] and Plauché et al. [1997] have suggested that the spectral peak (3–4 kHz) found in velars constitutes an 'all-or-none' perceptual cue. They propose that listeners are more likely to miss this cue than to introduce it spuriously. Results from Plauché et al. showed that when the spectral peak was filtered out, [ki] stimuli were heard as [ti] 100% of the time and [ti] was never heard as [ki].

It is not clear, however, if an all-or-none cue is at work in the data presented here. Careful examination of the long-term average spectra for [ki] and [tʃi] tokens revealed no systematic differences in overall spectral shape or location of peaks. Figure 12 gives a typical example of the long-term average spectra for [ki] and [tʃi] spoken by a single speaker in faster speech. Typically, both the [ki] and [tʃi] spectra have 2–3 peaks around 3–5 kHz, sharply drop to about 6 Hz, and then level out. However, the palatoalveolar affricate tokens (e.g., 12b) tend to have greater amplitude than the velar stops. Affricates are known to have more energy than stops, longer durations, and longer rise times (distance from consonant onset to energy peak) than stops [Shinn, 1978, and references therein]. Perhaps the greater energy and longer duration of the affricates create such a perceptually robust segment, that it is not likely to be misheard at all. In this study, [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 [Ahmed and Agrawl, 1969; Wang and Bilger, 1973].

In summary, velar stops before front vowels and palatoalveolar affricates were found to be acoustically similar, and more so for the voiceless than the voiced consonants. 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]. Thirdly, [tʃ] is rarely confused with [k]. These parallels argue strongly for an account of velar palatalization that involves a perceptual component.

Acknowledgements

Many thanks are due to the author's dissertation committee, Nicola Bessell, Andrew Garrett, Björn Lindblom, Scott Myers, Harvey Sussman, and Anthony Woodbury, as well as to the reviewers, Pat Keating and John Ohala, and the editor, Randy Diehl. This research has greatly benefited by their helpful comments and suggestions.

Appendix 1. Words Used in the Acoustic Study

	i	ɪ	eɪ	ɛ	æ	ɑ	ɔ	oʊ	ʊ	u	ʌ
g	geese	gift	gate	guess	gap	got	gawk	goat	guk	goose	gut
k	quiche	kick	cake	catch	cat	cot	cough	coat	cook	coop	cut
dʒ	jeep	Jif	Jake	jet	Jack	jot	josh ([ɑ])	joke	–	juice	jut
tʃ	chief	chip	chase	chess	chat	chop	chalk	choke	–	chew	chuck

Appendix 2. Questions for the Faster Speech Data

[g] and [dʒ]

Target word	Primary pair	Secondary pair	Target word	Primary pair	Secondary pair
geese	Canada geese Snow Geese	white geese colored geese	jeep	road jeep army jeep	4-wheel jeep in-town jeep
gift	gag gift real gift	thoughtful gift usual gift	Jif	creamy Jif crunchy Jif	regular Jif low-fat Jif
gate	decorative gate functional gate	fence gate wall gate	Jake	man named Jake dog named Jake	tall Jake short Jake
guess	wild guess educated guess	good guess bad guess	jet	aircraft jet whirlpool jet	water jet gas jet
gap	insignificant gap serious gap	bridging the gap widening the gap	Jack	red jack black jack	diamond jack spade jack
gut	beer gut fat gut	cat gut pork gut	jut	limestone jut granite jut	rocky jut snowy jut
got	being got getting got	somewhat got very got	jot	quick jot careless jot	writing jot typing jot
gawk	leering gawk dumbfounded gawk	surprised gawk curious gawk	josh	friendly josh reprimanding josh	playful josh meaningful josh
goat	Billy goat mountain goat	wild goat domesticated goat	joke	tasteless joke rude joke	funny joke stupid joke
guk [gʊk]	slimy guk sticky guk	green guk pink guk	* [dʒʊ]		
goose	barnyard goose wild goose	silly goose farm goose	juice	orange juice tomato juice	cranberry juice cranapple juice

Subjects were first asked to tell the difference between the primary pair. If four tokens of the target word were not obtained, the secondary pair was used. The secondary pair was only used four times.

[k] and [tʃ]

Target word	Primary pair	Secondary pair	Target word	Primary pair	Secondary pair
quiche	vegetable quiche ham quiche	cheese quiche mushroom quiche	chief	fire chief Indian chef	army chief navy chief
kick	karate kick regular kick	horse kick mule kick	chip	potato chip tortilla chip	salted chip flavored chip
cake	birthday cake wedding cake	chocolate cake strawberry cake	chase	high speed chase wild goose chase	dog chase cat chase
kept	being poorly kept being well kept	not being well kept being well kept	chess	tournament chess recreational chess	American chess Russian chess
cat	Siamese cat regular old cat	black cat calico cat	chat	chit chat serious chat	meaningful chat mindless chat
cut	prime cut choice cut	superficial cut serious cut	chuck	beef chuck ground chuck	drill chuck lathe chuck
cot	army cot nap cot	camping cot overnight cot	chop	pork chop lamb chop	karate chop regular chop
cough	hacking cough productive cough	dry cough persistent cough	chalk	blackboard chalk hand chalk	regular chalk colored chalk
coat	fall coat winter coat	pea coat army coat	choke	dog choke engine choke	sputtering choke coughing choke
cook	gourmet cook homestyle cook	fry cook line cook	* [tʃʊ]		
coop	chicken coop turkey coop	traditional coop industrial coop	chew	Redman chew mint chew	good chew bad chew

Discussion

The word used to elicit the [ki] sequence, *quiche,* ended in a palatoalveolar. It is conceivable that the [k] was affected by anticipatory coarticulation to the [ʃ]. In other words, perhaps there was some tongue-blade anticipation during the [k] that affected its acoustics. To investigate this possibility the peak spectral frequency of [ki] from *quiche* was compared to the peak spectral frequency of [ki] from words not containing a palatoalveolar.

More data was collected from 2 native English female speakers from Alabama. Using the faster speech methodology, 8 tokes of *quiche, key,* and *keep* (n.) were collected from each of the speakers. The final segments of the modifying words were kept constant across the 3 target words. Two questions were asked for each word. Each question elicited 4 tokens of each word. The questions are presented in the table below. Other questions were asked as warm-up and filler, but were not analyzed and are not presented here.

Target word	First question	Second question
quiche	vegetable quiche cheese quiche	farm quiche mushroom quiche
key	musical key house key	alarm key dustroom key
keep	castle keep house keep	farm keep jailroom keep

The target words were digitized and peak spectral frequency of the burst and aspiration of the [k] was measured as described in the methodology section of Experiment I. The peak frequencies were then submitted to a one-way ANOVA with repeated measures on word (*quiche, key* and *keep*). The ANOVA returned a nonsignificant finding [$F(2,30)=0.583$]. Thus, at least for the two speakers investigated here, the final palatoalveolar segment in *quiche* did not have a significant effect on the peak spectral frequency of the [k].

References

Ahmed, R.; Agrawl, S.: Significant features in the perception of (Hindi) consonants. J. acoust. Soc. Am. *45:* 758–763 (1969).

Anttila, R.; Historical and comparative linguistics (Benjamins, Philadelphia 1989).

Baxter, William, H.: A handbook of Old Chinese phonology (Mouton de Gruyter, New York 1992).

Behrens, S.; Blumstein S.: Acoustic characteristics of English voiceless fricatives: a descriptive analysis. J. Phonet. *16:* 295–298 (1988).

Bhat, D.: A general study of palatalization; in Greenberg, Universals of human language, vol 2: Phonology, pp. 47–92 (Stanford University Press, Stanford 1978).

Bloomfield, L.: Language (Holt, Rinehart, & Winston, New York 1993).

Campbell, A.: Old English grammar (Clarendon, Oxford 1959).

Chen, M.: Predictive power in phonological description. Lingua *32:* 173–191 (1973).

Cheng, C.-C.: Mandarin phonology; PhD diss., University of Illinois (1968).

Dorman, M.; Studdert-Kennedy, M.; Raphael, L.: Stop-consonant recognition: release bursts and formant transitions as functionally equivalent, context-dependent cues. Percept. Psychophys. *22:* 109–122 (1977).

Duez, D.: Second formant locus-nucleus patters in spontaneous speech: some preliminary results on French. Phonetic Experimental Research at the Institute of Linguistics University of Stockholm (PERILUS), vol. 10, pp. 109–114 (1989).

Duez, D.: Some evidence on second formant locus-nucleus patterns in spontaneous speech in French. Phonetic Experimental Research at the Institute of Linguistics University of Stockholm (PERILUS), vol. 11 (1990).

Emerson, O.: The history of the English language (MacMillian, New York 1903).

England, N.: El Mam: semejanzas y diferencias regionales; in England, Elliot, Lectura sobre la lingüística maya (CIRMA, La Antigua Guatemala 1990).

Fant, G.: Stops in CV-syllables; in Fant, Speech sounds and features, pp. 110–139 (MIT Press, Cambridge 1973).

Fischer-Jørgensen, E.: Acoustic analysis of stop consonants. Misc. Phonet. *2:* 42–59 (1954).

Fowler, C.: Invariants, specifier, cues: an investigation of locus equations as information for place of articulation. Percept. Psychophys. *55:* 597–610 (1994).

Grammont, M.: Traité de phonétique (Delagrave, Paris 1933).

Guion, S.: 1996. Velar palatalization: coarticulation, perception and sound change; PhD diss., University of Texas at Austin 1996.

Guthrie, M.: Comparative Bantu, vol. 1–4 (Gregg, London 1967–1970).

Halle, M.; Huges G.; Radley, J.-P.: Acoustic properties of stop consonants. J. acoust. Soc. Am. *21:* 107–116 (1957).

Harris, K.: Cues for the discrimination of American English fricatives in spoken syllables. Lang. Speech *1:* 1–7 (1958).

Heinz, J.; Stevens, K.: On the properties of voiceless fricative consonants. J. acoust. Soc. Am. *33:* 589–596 (1961).

Hock, H.: Principles of historical linguistics; 2nd ed. (Mouton de Gruyter, New York 1991).

Hoffmann, K.: Avestan language; in Yarshater, Encyclopaedia Iranica, pp. 47–62 (London 1982).
Hogg, R.: On English palatalization. Trans. Philol. Soc. 1979: 89–113 (1979).
Hughes, G.; Halle, M.: Spectral properties of fricative consonants. J. acoust. Soc. Am. *28:* 303–310 (1956).
Hyman, L.; Moxley, J.: The morpheme in phonological change: velar palatalization in Bantu. Diachronica *13:* 259–282 (1996).
Keating, P.; Phonetic and phonological representation of stop consonant voicing. Language *60:* 286–319 (1984).
Keating, P.; Lahiri, A.: Fronted velars, palatalized velars, and palatals. Phonetica *50:* 73–101 (1993).
Kewley-Port, D.: Time-varying features as correlates of place of articulation on stop consonants. J. acoust. Soc. Am. *73:* 322–355 (1983).
Kinkade, D.: The alveopalatal shift in Cowlitz Salish. Int. J. Am. Linguistics *39:* 224–231 (1973).
Kiparsky, V.: Russische historische Grammatik, vol. 1: die Entwicklung des Lautsystems (Winter, Heidelberg 1963).
Klatt, D.: Voice onset time, frication, and aspiration in word-initial consonant clusters. J. Speech Hear. Res. *18:* 686–706 (1975).
Kohler, K.J.: Segmental reduction in connected speech in German: phonological facts and phonetic explanations; in Hardcastle, Marchal, Speech production and speech modeling, pp. 69–92 (Kluwer, Dordrecht 1990).
Krull, D.: Acoustic properties as predictors of perceptual responses: a study of Swedish voiced stops. Phonetic Experimental Research at the Institute of Linguistics University of Stockholm (PERILUS), vol. 7 (1988).
Krull, D.: Second formant locus patterns and cosonant-vowel coarticulation in spontaneous speech. Phonetic Experimental Research at the Institute of Linguistics University of Stockholm (PERILUS), vol. 10, pp. 87–108 (1989).
Krull, D.: Locus-nucleus relation and VOT in spontaneous and elicited speech. Proc. 12th Int. Congr. Phonet. Sci., vol. 4, pp. 382–385 (1991).
Lehiste, I.; Peterson, G.: Vowel amplitude and phonemic stress in American English. J. acoust. Soc. Am. *31:* 428–435 (1959).
Lindblom, B.: On vowel reduction, Report 29 (Spreech Transmission Laboratory, Stockholm 1963).
Lisker, L.; Abramson, A.: A cross-language study of voicing in initial stops: acoustical measurements. Word *20:* 384–422 (1964).
Lunt, H.: Old Church Slavonic grammar (Mouton, The Hague 1974).
Matthews, W.: Russian historical grammar (Athlone, London 1967).
Mayrhofer, M.: Sanskrit Grammatik mit sprachvergleichenden Erläuterungen; 2nd ed. (Berlin, 1965). English translation Ford, G.: A Sanskrit Grammar (University of Alabama Press 1972).
Myers, S.: Epenthesis, mutation, and structure preservation in the Shona causative. Stud Afr Linguistics *23:* 185–216 (1992–1994).
Nartey, J.: On fricative phones and phonemes: measuring the phonetic differences within and between languages. UCLA Working Papers in Phonetics, No. 55 (1982).
Neeld, R.: Remarks on palatalization. Working Papers in Linguistics No. 14: Studies in phonology and methodology (Columbus, Department of Linguistics, The Ohio State University 1973).
Ohala, J.: The role of physiological and acoustic models in explaining the direction of sound change. Project on linguistic analysis. Reports, Second Series, No. 15. Berkeley: Phonology Laboratory – Department of Linguistics, University of California, Berkeley 1971).
Ohala, J.: Phonetic explanation in phonology; in Bruck, Fox, LaGaly, Papers from the parasession on natural phonology, pp. 251–274 (Chicago Linguistic Society, Chicago 1974).
Ohala, J.: The listener as the source of sound change; in Masek, Hendrick, Miller, Papers from the parasession on language and behavior, pp. 178–203 (Chicago Linguistic Society, Chicago 1981).
Ohala, J.: Linguistics and automatic speech processing, in Mori, Suen, New systems and architectures for automatic speech recognition and synthesis. NATO ASI Series, Series F: Computer and Systems Sciences, vol. 16, pp. 447–475 (Springer, Berlin 1985).
Ohala, J.: Discussion, in Perkell, Klatt, Invariance and variability in speech processes, pp. 197–198 (Erlbaum, Hillsdale 1986).
Ohala, J.: What's cognitive, what's not, in sound change; in Kellermann, Morrissey, Diachrony within synchrony: language history and cognition. Duisburger Arbeiten zur Sprach- und Kulturwissenschaft 14, pp. 309–355 (Lang, Frankfurt 1992).
Ohala, J.: The origin of sound patterns in vocal tract constraints; in MacNeilage, The production of speech, pp. 189–216 (Springer, New York 1983).
Ohala, J.: Sound change is drawn from a pool of synchronic variation; in Breivik, Jahr, Language change: contributions to the study of its causes, pp. 171–198 (Mouton de Gruyter, New York 1989).
Ohala, J.: The phonetics of sound change; in Jones, Historical linguistics: problems and perspectives, pp. 237–278 (Longman, London 1993).
Peterson, G.; Barney, H.: Control methods used in a study of the vowels. J. acoust. Soc. Am. *24:* 175–184 (1952).
Plauché, M.; Delogu, C.; Ohala, J.: Asymmetries in consonant confusion. Eurospeech 97 (in press).
Poch-Olivé, D.; Fernandez-Guitierrez, N.; Martinez-Dauden, G.: Some problems of coarticulation in CV stop syllables in Spanish and Catalan spontaneous speech. Proc. Speech Res. '89, June 1–3. Budapest, pp. 111–115 (1989).
Pulleyblank, E.: Middle Chinese: a study in historical phonology (University of British Columbia Press, Vancouver 1984).

Repp, B.; Lin, H.-B.: Acoustic properties and perception of stop consonant release transients. J. acoust. Soc. Am. *85:* 379–395 (1989).
Schenker, A.: Proto-Slavic; in Comrie, Corbett, The Slavonic languages (Routledge, New York 1993).
Shevelov, G.: A prehistory of slavic (Columbia University Press, New York 1965).
Smiths, R.: Detailed versus gross spectro-temporal cues for the perception of stop consonants; PhD diss., Technische Universiteit Eindhoven (1995).
Smits, R.; ten Bosch, L.; Collier, R.: Evaluation of various sets of acoustic cues for the perception of prevocalic stop consonants. I. Perception experiment. J. acoust. Soc. Am. *100:* 3852–3864 (1996).
Strevens, P.: Spectra of fricative noise in human speech. Lang. Speech *3:* 32–49 (1960).
Sussman, H.; Fruchter, D.; Cable, A.: Locus equations derived from compensatory articulation. J. acoust. Soc. Am. *97:* 3112–3124 (1995).
Sussman, H.; Hoemeke, K.; Ahmed, F.: A cross-linguistic investigation of locus equations as a phonetic descriptor for place of articulation. J. acoust. Soc. Am. *94:* 1256–1268 (1993).
Sussman, H.; McCaffery, H.; Matthews, S.: An investigation of locus equations as a source of relational invariance for stop place categorization. J. acoust. Soc. Am. *90:* 1309–1325 (1991).
Sussman, H.; Shore, J.: Locus equations as phonetic descriptors of consonantal place of articulation. Percept. Psychophys. *58:* 936–946 (1996).
Tekieli, M.; Cullinan, W.: The perception of temporally segmented vowels and consonant-vowel syllables. J. Speech Hear. Res. *22:* 103–121 (1979).
Vaillant, A.: Grammaire comparée des langues slaves, vol. 1: phonétique (I A C., Lyon 1950).
Wang, M.; Bilger, R.: Consonant confusions in noise: a study of perceptual features. J. acoust. Soc. Am. *54:* 1248–1266 (1973).
Whalen, D.: Perception of the English /s/–/ʃ/ distinction relies on fricative noises and transitions, not on brief spectral slices. J. acoust. Soc. Am. *90:* 1776–1785 (1991).
Winitz, H.; Scheib, M.; Reeds, J.: Identification of stops and vowels for the burst portion of /p, t, k/ isolated from conversational speech. J. acoust. Soc. Am. *51:* 1309–1317 (1972).
Zue, V.: Acoustic characteristics of stop consonants: a controlled study. Technical Report 523 (Lincoln Laboratory, Cambridge 1976).