CHAPTER

Markedness and Consonant Confusion Asymmetries

Steve S. Chang, Madelaine C. Plauché, and John J. Ohala

Department of Linguistics
University of California, Berkeley
Berkeley, California 94720

The Role of Speech Perception in Phonology

1. INTRODUCTION

In both lab-based speech perception studies and in sound change, one finds directional asymmetries of consonant change. In each of the following examples, the CV sequences on the left have shifted diachronically to those on the right, but the reverse is rarely attested:

ki	>	tʃi	(e.g., Slavic, Indo-Iranian, Bantu)[1]
pi	>	ti	(cf. Czech dial. var: pǐ:vo [pʲi:vɔ] ~ [ti:vɔ] 'beer')
ku	>	pu	(cf. PIE ek^{w}ōs 'horse' Gk hippos).

For insight into this phenomenon, we turn to the domain of vision. Below are examples of unidirectional confusions made by subjects when asked to identify Roman capitalized letters in a visual perception study conducted by Gilmore *et al.* (1979):

E	**>**	**F**	**Q**	**>**	**O**
R	**>**	**P**	**W**	**>**	**V**[2]

In visual perception tasks, subjects confused 'E' with 'F,' 'R' with 'P,' and so on, but 'F' was rarely confused with 'E' and 'P' rarely with 'R,' etc. In each pair of confused characters, both are structurally similar, but the one on the left has an 'extra' feature that the one on the right lacks (Gilmore *et al.*, 1979). The perceiver is more likely to miss a distinctive part of the stimulus array than she is to imagine its presence when it is not actually there.

The same conceptual explanation may be applied to consonant confusion asymmetries. In speech, the pertinent cues for differentiation are the temporal and spectral properties of acoustic events. These acoustic cues have intrinsically different salience or robustness for the task of differentiation. For example, acoustic cues from the amplitude envelope are more auditorily salient than those from spectral details (Miller & Nicely, 1954). Irrespective of intrinsic salience, all cues are subject — probabilistically — to degradation in transmission. The second law of thermodynamics dictates an increase in entropy; entropy never decreases (absent the diversion of energy from other domains). While all cues are subject to degradation, robust cues are less likely to be compromised to the point of losing their distinctiveness. Non-robust cues, on the other hand, may be degraded to the point where they lose their distinctiveness. Significant degradation of a non-robust cue can lead to an irreversible confusion. We argue that it is precisely such an instance that gives rise to asymmetries in confusion matrices.

It has been observed that [ki] is often confused in a laboratory setting for [ti] by listeners, but [ti] is almost never confused as [ki] (Winitz *et al.*, 1972; Delogu *et al.*, 1995). Plauché *et al.* (1997) investigated the temporal and spectral properties responsible for this asymmetrical confusion in stop place. They found

that, due to the resulting raised F2 of the following high vowel, the formant transitions for [ki] and [ti] are similar, neutralizing the formants' role as distinctive stop place cues. The spectra of the stop bursts, too, are similar, but the stop burst in [ki] has an "extra," non-robust feature: a compact mid-frequency spectral peak — essentially the front cavity resonance ≡F3 (Stevens & Blumstein, 1978). Plauché *et al.* (1997) hypothesized that this mid-frequency spectral peak functions much like the extra "foot" in the letter 'E' or the extra "tail" in the letter 'Q.' If this "extra," non-robust feature is degraded to the point of losing its contrastiveness, listeners are likely to confuse the [ki] as [ti]. [ti] will very rarely, however, be confused as [ki] since listeners are unlikely to erroneously insert a nonexistent cue into the speech signal for [ti]. This accounts for the asymmetry.

By filtering out the characteristic mid-frequency spectral peak of the velar burst in [ki], Plauché *et al.* (1997) succeeded in enhancing the asymmetrical confusion with [ti]. Table 4.1, adapted from Plauché *et al.* (1997), summarizes the results from a perceptual experiment in which English-speaking subjects were asked to identify the stop in three natural speech tokens and an edited token. The "edited ki" token was a natural [ki] whose stop burst had been band-reject filtered to effectively remove the characteristic mid-frequency peak of the velar burst.

TABLE 4.1

Confusion Asymmetry Results from Plauché *et al.* (1997)

Stimuli/ response	pi	ti	ki
pi	**97%**	3%	0
ti	0	**100%**	0
ki	0	20%	**80%**
Edited ki	0	**100%**	0

As the top half of the table demonstrates, there was some asymmetric confusion evident in the natural stimuli: 3% of the naturally produced [pi] tokens and 20% of the naturally produced [ki] tokens were mistaken for [ti]. All of the [ti] tokens, however, were correctly identified. The filtered [ki] token, "edited ki," was always identified as [ti].

Two important questions were raised with respect to the results of the Plauché *et al.* study. First, how do we know that the acoustic properties of [ki] and [ti] (e.g., the "extra" feature consisting of the mid-frequency peak) are responsible

for the consonant confusion asymmetry and not markedness? After all, alveolar stops occur with much greater frequency than velar stops in the languages of the world, and velar stops are known in general to be more marked than alveolar stops. Lyublinskaya (1966), for example, argued that frequency effects were present even at the level of phoneme identification. Aren't these sound changes and confusion asymmetries simply a favoring of the unmarked? Second, the [ki] > [ti] confusion asymmetry displayed in the laboratory is rarely attested in actual diachronic sound change. [ki] is known to change to [tʃi] in many languages, not to [ti]. Given that [ki] is so readily mistaken for [ti] in the laboratory, that [ki] > [ti] is not a more frequent sound change poses an interesting question. Why don't subjects in laboratory perception studies hear [tʃi] instead of [ti]? This study addresses both of these questions.

Study 1 replicates and expands Plauché *et al.* (1997) and demonstrates the primacy of the acoustic/auditory characteristics over markedness as an explanation for this consonant confusion asymmetry by inducing confusion in the opposite direction ([ti] > [ki]). Study 2 provides further support for the primacy of auditory cues over markedness by illustrating that stop place confusion is specific to vocalic environments in which all but one differentiating cue is neutralized, namely, the mid-frequency peak. Finally, studies 3a and 3b address the question of the apparent discrepancy between the actual ki > tʃi sound change and the [ki] > [ti] confusion asymmetry induced in the laboratory.

II. MARKEDNESS

Before addressing the role that markedness might play in asymmetrical consonant confusion, it is important to define what we mean by "markedness." The weak version of our argument would define markedness as an effect due to crosslinguistic, universal frequency of occurrence. That is, due to the relatively high frequencies with which certain sounds occur in the languages of the world, in perception, where some ambiguity may arise, such sounds are the "default" percept. Alveolar stops are more frequent and therefore less marked than velar stops. Conceivably, this fact may bias listeners to think they hear alveolar stops when presented with degraded velar stops.

The stronger version of our argument defines markedness as any non-acoustic considerations that listeners employ in the task of differentiation. Such considerations may be SPE-style (Chomsky & Halle, 1968) universal disfavoring of certain sounds based either on articulatory or perceptual (but non-acoustic) ease, a Praguean positive valence for a particular feature (Trubetzkoy, 1939), or any other traditional definitions of markedness. Under this definition of markedness, velar stops may be argued as being universally more marked than alveolar stops due to any number of reasons, for example, articulatory or perceptual.

While we acknowledge that frequency of occurrence and other universal markedness constraints may indeed contribute to consonant confusion asymmetries, we believe they are secondary and may effectively be factored out. Study 1 (§II.A) supports the explanatory primacy of acoustic cues over markedness factors in confusion asymmetries by inducing confusion from a less marked to a more marked consonant ([ti] > [ki]) by manipulating the acoustic features of the bursts. Study 2 (§II.B) further supports this hypothesis by demonstrating that the consonant confusion asymmetry is obtained only in those vocalic contexts, namely, high front vowels, where all other velar cues except the mid-frequency spectral peak are neutralized.

A. Study 1

1. Stimuli Preparation

For Study 1, three male native speakers of American English in their mid to late twenties (BB, SC, and JR) were recorded to create the test tokens. In order to obtain voiceless, unaspirated stops, the target CV tokens [ki], [pi], and [ti] were digitally extracted from the words 'skeet,' 'speak,' and 'steep,' that is, by removing the [s] and end-truncating the vowel. Voiceless, unaspirated stops were selected to (1) simulate the Spanish CV tokens in the Plauché *et al.* study and (2) avoid the messiness of editing and identifying the components of the aperiodic noise portion characteristic of English aspirated stops. The target words were read five times by each speaker in the carrier sentence 'Say ___ once.' The tokens were digitally captured using Kay Elemetrics Corporation's Computerized Speech Laboratory (CSL) with a sampling rate of 20,000.

For each subject, one [pi], one [ti], and one [ki] token, all with similar prosody, were selected. The Plauché *et al.* (1997) study filtered the characteristic velar burst of [ki] with a Hanning band-reject filter of order 10 between the frequencies 2.5 and 4.0 kHz. Here we adopt the same methodology but use four different Blackman filters designed from combinations of two orders (**L**ow/**H**igh) and two bandwidths (**N**arrow/**W**ide). The purpose of using four different filters was to gauge the degree of degradation necessary to induce confusion asymmetries. One hundred percent of the subjects in the Plauché *et al.* study perceived the filtered token as being [ti] (Table 4.2). We wanted to demonstrate that it is indeed the "extra" feature of the mid-frequency spectral burst in [ki] that differentiates it from a [ti] by showing a graded effect on perception by degrading the spectral peak to differing degrees.

TABLE 4.2

Filter Parameters for Study 1

Filter parameters	Low order (20)	High order (35)
Narrow bandwidth (1 kHz)	KILN	KIHN
Wide bandwidth (2 kHz)	KILW	KIHW

The filters were centered at each speaker's velar mid-frequency peak. BB = 3370 Hz, SC = 3390 Hz, and JR = 2751 Hz.

In addition to the 12 filtered [ki] stimuli (4 filtered [ki] tokens per speaker), in which we removed the characteristic velar mid-frequency spectral peak, we processed 3 [ti] tokens (1 mixed [ti] token per speaker) to introduce the mid-frequency peak to the [ti] burst. In order to add energy to the mid-frequency region of the alveolar stop in [ti], we (1) generated white noise (sampling rate 20,000; duration: 48 ms), (2) band-pass filtered the white noise, using a Blackman window of order 101 from 2880 to 3880 Hz, which corresponds to the average center

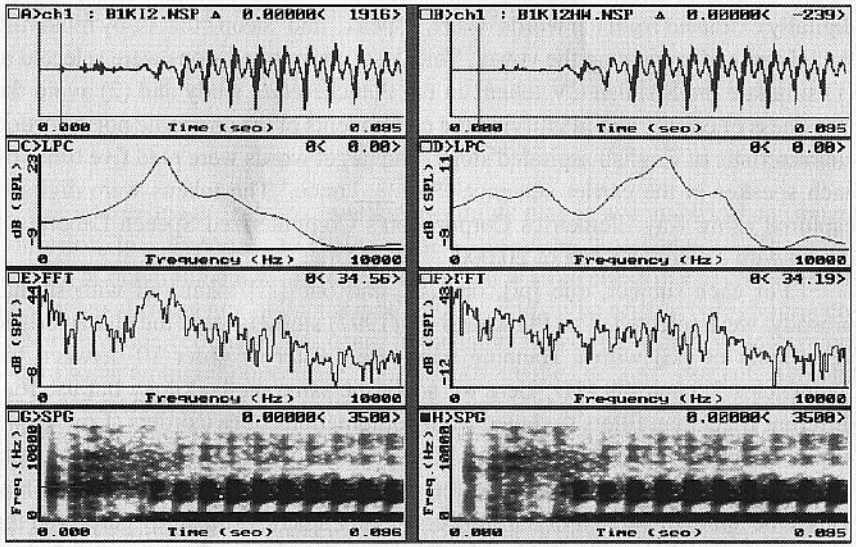

Figure 4.1. Waveform, LPC at burst, FFT at burst, and spectrogram of [ki] and KIHW. Note the characteristic mid-frequency spectral peak of the velar burst in the left column. The spectral peak in the right column has been attenuated by filtering. (Note that the spectral scales have not been normalized.)

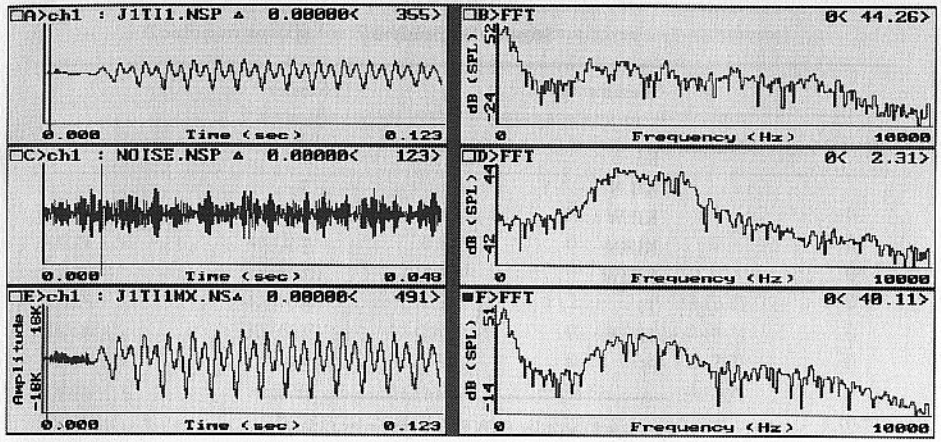

Figure 4.2. Waveform and FFT at burst of [ti] (top panel), filtered white noise (middle), and mixed [ti] (TIMX). The spectral shape of the filtered white noise, when mixed with a [ti] burst, yields a mid-frequency spectral peak akin to that found in a [ki] burst (compare with left FFT panel of Figure 4.1).

frequency of the spectral peak of [ki] across all three speakers, and (3) mixed the filtered noise with the burst of the alveolar stop in [ti].

The purpose of filtering the mid-frequency peak out of the velar bursts, as in Plauché *et al.* (1997), is to demonstrate that the spectral peak functions as an "extra" feature, potentially causing asymmetrical confusion. By introducing this peak into an alveolar stop burst, we hypothesize that some listeners will confuse the alveolar stop with a velar stop, which cannot be accounted for with a "default" markedness approach.

2. Perception Experiment

For distributional purposes, we added three additional [ti] tokens and nine additional [pi] tokens. The number of stimuli for Study 1 are listed in Table 4.3.

Sixteen speakers of American English, 12 women and 4 men, ranging in age from 19 to 33, served as subjects for the three studies. All 16 subjects were students at the University of California, Berkeley, at the time of the studies and were paid for their participation. None of the subjects had any known hearing conditions that would have affected their performance. However, one male subject's results were excluded as investigation of his data showed that he did not understand the task. The stimuli were randomized and presented over headphones in a quiet environ-

TABLE 4.3
Stimuli for Study 1

Stimulus	Number
KI	3
KILN	3
KILW	3
KIHN	3
KIHW	3
TI	6
TIMX	3
PI	12

L = low order (20), H = high order (35), N = narrow bandwidth (1000 Hz), and W = wide bandwidth (2000 Hz).

ment using Kay Elemetrics Corporation's Auditory Perception Program and Database (ASPP).

Subjects were asked to identify the consonant in each stimulus, from the choices B/P, D/T, CH, and G/K. (We were concerned with the perceived place of articulation, not the voicing of the segments in question, especially since the tokens had VOT values typical of English word-initial voiced stops but were phonemically voiceless, e.g., *skeet*.) Although the responses were not timed, the subjects were instructed to answer as quickly as possible and were not allowed to change their answers.

3. Results

The confusion matrix for the listeners' responses of Study 1 is presented in Table 4.4. Unlike the subjects from the Plauché *et al.* (1997) study, who confused natural [ki] as a [ti] 20% of the time, the subjects in the current study correctly identified 100% of the natural [ki] tokens. As expected, filtering the burst of the [ki] tokens did induce confusion with [ti]. It is interesting to note, however, that only the filter with the high order and wide bandwidth yielded significant confusion. The other three filters had a negligible effect, if any, on perception of the velar stop. This indicates that the degradation threshold at which listeners confuse [ki] for [ti] corresponds to a band-pass filter with an order of 35 and a bandwidth of 2 kHz centered around 3000 Hz. It is also worth noting that this is the first forced-choice stop place confusion study to offer both [ti] and [tʃi] as possible responses for

TABLE 4.4
Confusion matrix for Study 1 (percentages are given in parentheses)

	Response				
Stimulus	G/K (%)	D/T (%)	B/P (%)	CH (%)	None (%)
KI	45 (100)	0	0	0	0
KILN	45 (100)	0	0	0	0
KILW	43 (95.6)	2 (4.4)	0	0	0
KIHN	42 (93.3)	1 (2.2)	1 (2.2)	1 (2.2)	0
KIHW	30 (66.7)	13 (28.9)	0	2 (4.4)	0
TI	0	87 (97.8)	0	2 (2.3)	0
TIMX	4 (8.9)	31 (68.9)	6 (13.3)	4 (8.9)	0
PI	0	4 (2.2)	175 (97.2)	0	1 (0.6)

processed [ki] stimuli. Most previous consonant confusion studies have been forced-choice tasks that did not offer subjects the option of selecting either [t] or [tʃ] as the percept (e.g., Plauché *et al.*, 1997; Miller & Nicely, 1955; Winitz *et al.*, 1972; Wang & Bilger, 1973; Delogu *et al.*, 1995). Guion (1998), on the other hand, only offered [k], [tʃ], [g], and [dʒ], to the exclusion of [t] and [d].

Upon further investigation, we discovered that there was a strong speaker-dependent effect, suggesting that other cues are involved in identification of stop place. Speaker BB's stimuli were responsible for the majority (12 out of 13) of the listeners' confusion of KIHW for [ti]. An examination of burst characteristics by speaker revealed that BB's velar stop token had the shortest VOT, shortest burst duration (onset of the first burst to the end of the final burst), and only three bursts (Table 4.5). We suspect that the relatively long VOT, long burst duration, and

TABLE 4.5
Confusion of KIHW by Speaker

	KIHW					
	Response			Velar characteristics		
Speaker	G/K	D/T	CH	VOT	burst	# of bursts
BB	2	12	1	26 ms	15 ms	3 bursts
SC	14	1	0	38 ms	20 ms	3 bursts
JR	14	0	1	60 ms	29 ms	4 bursts

multiple number of bursts of SC and JR's tokens provided sufficient cues for listeners to detect velarity, even after filtering with a high order and broad bandwidth. When BB's mid-frequency spectral peak is effectively filtered out, on the other hand, VOT, burst duration and the number of bursts did not provide sufficient cues for velar place.

TABLE 4.6
Confusion of TIMX by Speaker

	TIMX					
	Response				Alveolar characteristics	
Speaker	G/K	D/T	B/P	CH	VOT	# of bursts
BB	0	14	0	1	13 ms	1
SC	3	9	0	3	37 ms	1
JR	1	8	6	0	16 ms	1

The processed [ti] file, in which we introduced a mid-frequency spectral peak to the alveolar burst, induced confusion as well. However, directionality toward [ki] was not evident. It would appear that for more than 30% of the TIMX tokens subjects were aware that the burst in TIMX was not alveolar, but they were unsure about the identity of the stop place, as these tokens were equally confused as [ki], [pi], and [tʃi].

These results, too, reflect some speaker-dependent effects, as summarized in Table 4.6. Speaker BB's token did not induce significant confusion, and speakers SC and JR's tokens resulted in different patterns of confusion. At least 93% (28 out of 30) of each speaker's natural [ti] tokens were correctly identified as alveolar. Three of SC's tokens were confused as velars, and three others were confused as [tʃ]. We suspect that the latter is due in part to the relatively long VOT of this token. Nearly half of JR's tokens were confused as bilabials, a result that we were not anticipating. Speaker SC's tokens best support the hypothesis adopted in this paper. Mixing in the mid-frequency spectral peak characteristic of velar stops to an alveolar burst did induce some listeners to perceive a velar stop. Moreover, some listeners perceived [tʃ], indicating that [ti] and [tʃi] are, in fact, perceptually linked, a point to which we will return in Study 3

The results from Study 1 support the view that listeners are opportunists. They make use of whatever cues are available for differentiating speech sounds.

Even when the characteristic "extra" feature of the mid-frequency spectral peak of a velar stop is removed, the presence of other velar cues may be sufficient for identification of the velar place. A further investigation of the relative salience of other acoustic cues such as VOT and the number of bursts, among others, is necessary for a fuller understanding of these phenomena. Language-specific considerations may also be important factors. In American English, for example, postalveolar fricatives and affricates are usually rounded, thus having lower center frequencies of the noise spectra than were presented in our stimuli.

B. Study 2

In Study 2 we seek to demonstrate that the [k] > [t] confusion asymmetry is specific to a high, front vocalic environment, further supporting the claim that it is the acoustic properties of [ki] and [ti] that are responsible for the consonant confusion asymmetry. A simple markedness account would be unable to explain why the confusion asymmetry is not obtained in all vocalic environments since velar stops are presumably more marked than alveolar stops regardless of the quality of the following vowel.

1. Stimuli Preparation

For Study 2, two of the speakers from Study 1, BB and SC, and a third subject, MP, a female native speaker of American English in her mid-twenties, were recorded to create the test stimuli. The target tokens for Study 2 were unaspirated, voiceless kV tokens where the vowel was one of nine American English vowels (see Table 4.7). The tokens were again digitally extracted from sC clusters in the carrier sentence "Say ___ once." For control purposes, we also elicited tV tokens in which the consonant was a voiceless, unaspirated alveolar stop. The tokens were digitally captured using Computerized Speech Laboratory (CSL) at a sampling rate of 20,000.

For each subject, one representative token from the three repetitions was selected for prosodic consistency. The burst of each token was filtered to remove the mid-frequency spectral peak. LPC and FFT spectral analyses showed that the center peak frequency for the velar burst in front of these different vowels varied little for a given speaker. Consequently, one filter was designed for all the stimuli (order 35, bandwidth 2 kHz).

The same 16 subjects from Study 1 participated in Study 2 (the data of all 16 subjects were used for this study). The subjects were asked to identify the stop in the stimuli. They were given three choices: G/K, D/T, or CH. (Again, we were concerned with the perceived place of articulation, not the voicing of the segments

TABLE 4.7
Test Stimuli for Study 2

Velar	Filtered	Alveolar
ki	ki – flt	ti
kau	kau – flt	tau
ker	ker – flt	ter
ku	ku – flt	tu
kei	kei – flt	tei
kaɪ	kaɪ – flt	taɪ
kae	kae – flt	tae
ka	ka – flt	ta
kou	kou – flt	tou

in question.) Response time was not measured, but the subjects were instructed to answer as quickly as possible and were not allowed to change their answers. Given that all the distinctive acoustic cues for velar place (except the mid-frequency spectral peak of the burst) are neutralized only before high front vowels, we predicted that confusion would not be induced by filtering velar bursts before non-high, non-front vowels.

2. *Results*

The results for the kV tokens are summarized in Figure 4.3. As the vast majority of the stimuli were correctly identified as velars (see the confusion matrix in Table 4.8), only the confused responses are displayed. The histogram shows that stimuli [ki], for example, induced one CH response, indicating that the remaining 47 tokens were correctly identified as velars.

As expected, the filtered [ki] file resulted in the greatest number of D/T responses. Given that it is the acoustic properties of high front vowels that give rise to consonant confusion asymmetry, it is not surprising that we get a confusion effect for the other two filtered files with front vowels: filtered [keɪ] and filtered [kæ]. A surprising result was that unfiltered [kaɪ] elicited as many D/T responses as the front vowel filtered files. Moreover, this confusion disappeared when the [kaɪ] file was filtered. Another interesting result is that some of the stimuli elicited CH responses even though none of the stimuli were heavily aspirated. Stimuli that elicited multiple CH responses were those with relatively "tight constrictions," that is, tokens with relatively high vowels such as [ki] and [ku], or in the case of

TABLE 4.8
Confusion Matrix for Natural and Filtered kV Stimuli (where V is one of nine American English vowels)

	Response			
Stimulus	G/K	D/T	CH	B/P
KA	48	0	0	0
ka–flt	48	0	0	0
kae	46	1	1	0
kae–flt	44	4	0	0
kai	44	4	0	0
kai–flt	48	0	0	0
kau	48	0	0	0
kau–flt	45	2	1	0
kei	48	0	0	0
kei–flt	44	4	0	0
ker	46	0	2	0
ker–flt	45	0	3	0
ki	47	0	1	0
ki–flt	41	5	2	0
kou	46	1	1	0
kou–flt	48	0	0	0
ku	46	0	2	0
ku–flt	48	0	0	0

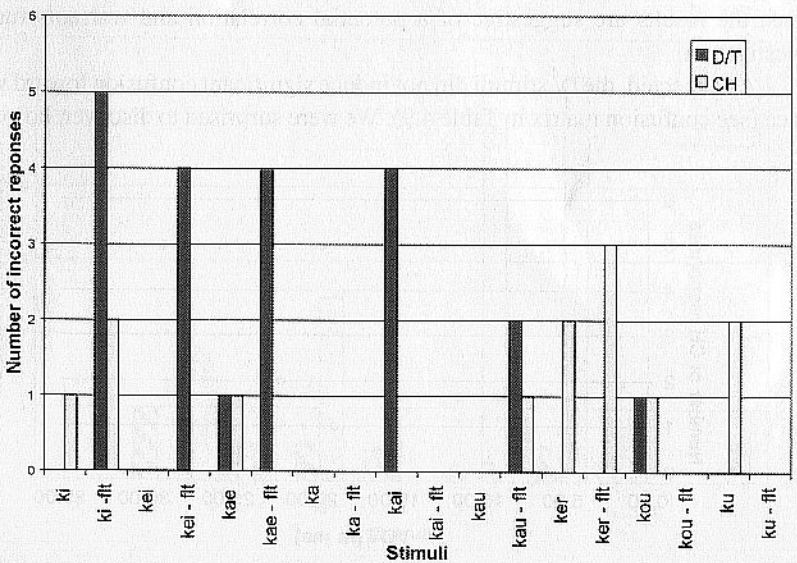

Figure 4.3. Confusion histogram for natural and filtered kV stimuli, where V is one of nine American English vowels. Only incorrect responses are shown.

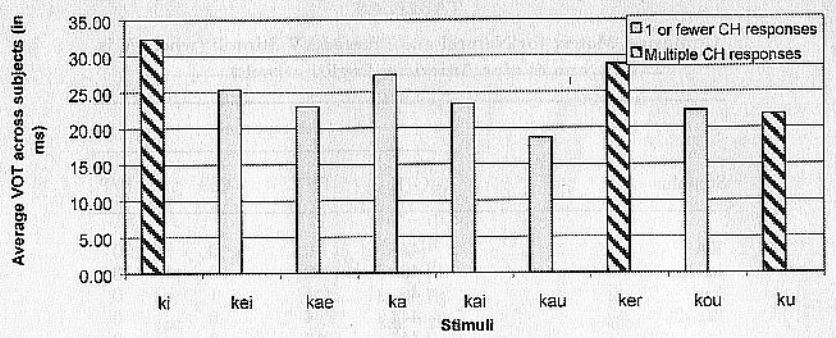

Figure 4.4. Average VOT across all three subjects of kV tokens. Those tokens that elicited multiple [tʃ] responses are cross-hatched.

[kɔ˞], a token with multiple constrictions (Chang, 1999) (see Figure 4.4). We hypothesize that the relatively tight constrictions of these vowels resulted in slightly longer VOT (as in [ki] and [kɔ˞]) or slight frication due to aerodynamic principles (Chang, 1999), leading some listeners to perceive a change in manner as well as place of articulation. Figure 4.5 charts the correlation between VOT (along the abscissa) and the number of times a token was perceived as being [tʃ] (along the ordinate). While there aren't enough data points to draw any conclusions, the results are suggestive of a potential correlation and warrants further investigation.

As expected, the tV stimuli did not induce significant confusion toward velar place (see confusion matrix in Table 4.9). We were surprised to discover, however,

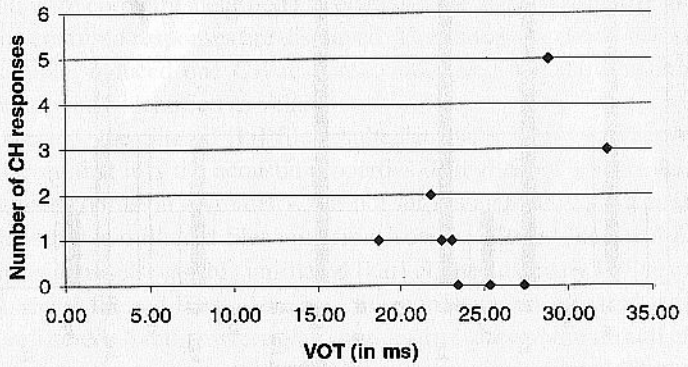

Figure 4.5. Correlation between VOT (averaged across all three subjects for kV tokens) and number of CH responses.

TABLE 4.9

Confusion Matrix for tV stimuli (where V is one of nine American English vowels)

	Response			
	G/K	D/T	CH	B/P
TA	1	42	5	0
TAE	0	47	1	0
TAU	0	38	10	0
TAI	1	47	0	0
TEI	2	46	0	0
TER	1	36	11	0
TI	0	15	0	0
TOU	0	48	0	0
TU	0	46	2	0

that four of the tV stimuli — [tɑ], [taʊ], [tu], and [tɔ˞] — induced multiple CH responses, as reflected in the confusion histogram in Figure 4.6, where only incorrect responses are shown. Closer investigation revealed that most of these

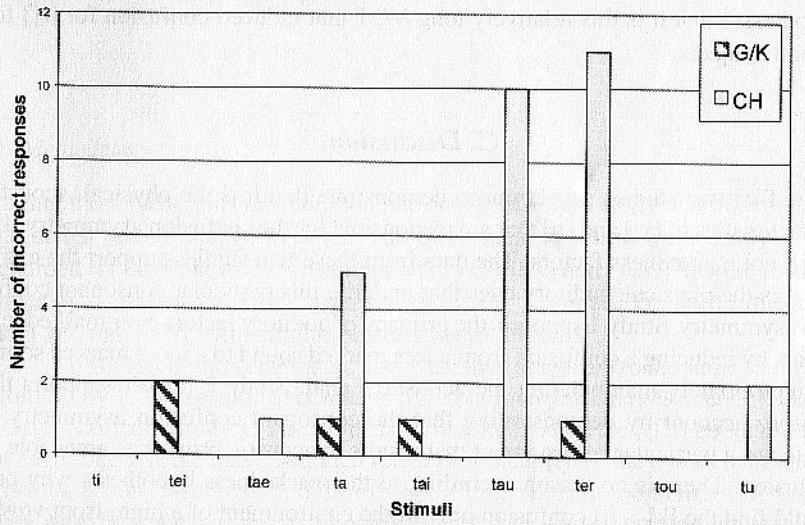

Figure 4.6. Confusion histogram from tV stimuli, where V is one of nine American English vowels. Only incorrect responses are shown.

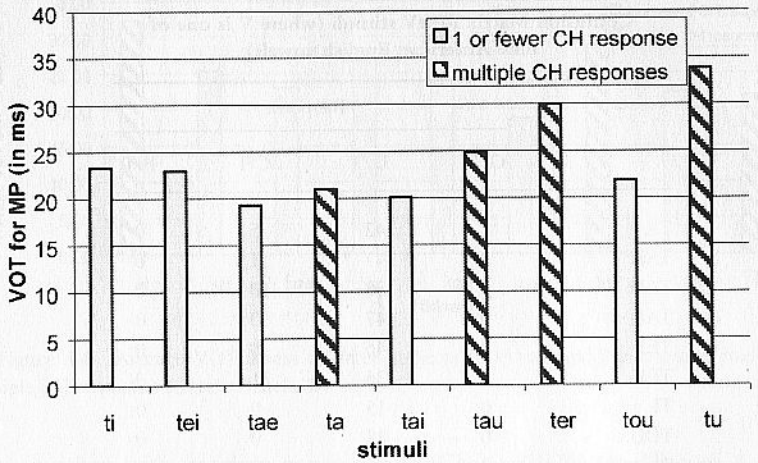

Figure 4.7. VOT measurements for subject MP. Dark bars indicate tokens that elicited mulitple (tʃ) responses.

confusions were for tokens uttered by the same subject, MP. A histogram of VOT for MP's tV tokens (Figure 4.7) reveals that, unlike subjects BB and SC, MP's utterances of [tɑ], [taʊ], [tu], and [tɚ] had relatively high VOT values. We hypothesize that it is this relatively long VOT that induced confusion for [tʃ] for several subjects.

C. Discussion

In the first two studies we set out to demonstrate that it is the physical, acoustic characteristics of [ki] and [ti] that are responsible for the confusion asymmetry [ki] > [ti], not markedness factors. The data from these two studies support the claim that it is the physical, auditory cues that underlie this particular consonant confusion asymmetry. Study 1 supports the primacy of auditory factors over markedness factors by inducing a confusion from a less marked sound to a more marked sound by appropriately manipulating the acoustic signal. Study 2 further supports the auditory account by demonstrating that the consonant confusion asymmetry is limited to a particular environment that results in acoustic properties amenable to confusion. There is no reason according to the markedness hypothesis why one should find the [k] > [t] confusion only in the environment of a high, front vowel since velar stops are presumably more marked than alveolar stops in all vocalic environments.

III. [ti] VERSUS [tʃi]

One potential criticism of Studies 1 and 2 is that ki > ti is not a common sound change. Rather, ki ubiquitously changes to tʃi or ts in the languages of the world, not to ti.

We hypothesize that the lack of a [ki] > [tʃi] confusion asymmetry in the laboratory is due to the combination of two facts about most laboratory studies of this sort. First, as mentioned earlier, previous consonant confusion asymmetry studies that deal with [ki] > [ti] did not offer both [ti] and [tʃi] as options. Second, the stops used in many laboratory studies (including this one) are unaspirated. It is unlikely, we think, that subjects will perceive a change in the manner of articulation as well as a change in the place of articulation for unaspirated stops. We suspect that it is the aspiration noise following the degraded burst of a [kʰi] that is reanalyzed as the [ʃ] of a [tʃ], that is, the listener "hears" an alveolar burst (as demonstrated by the first study) followed by an unusual, incongruous amount of aspiration for an alveolar stop (since velars are known to engender the longest VOT, especially before the high, front vowel [i]; Cho & Ladefoged, 1999). Faced with such acoustic cues, we argue, some listeners parse the token as [tʃi].

As mentioned above, it is also worth noting that language-specific considerations may play a role. For example, American English postalveolar fricatives and affricates are usually rounded. The relatively high F3 of the filtered file may be enough of a cue for American English subjects that the stop in the filtered [ki] token is not a postalveolar segment.

A. Study 3A

1. Methodology

Taking these facts into consideration, we designed a third study that better reflects the actual sound change ki > tʃi. The three male speakers from Study 1 were recorded for the final study. The target words *key* [kʰi] and *chi* [tʃʰi] were read five times by each speaker in the carrier sentence 'Say ___ once.' For each subject, one [kʰi] and one [tʃʰi], with similar prosody, were selected.

The bursts of the velar token, *key*, were then subjected to the same four filters from Study 1. The resulting six tokens were randomized and presented to the same 16 subjects from Study 1. The subjects were asked to rate the [tʃʰi] goodness score of each token. If the token sounded like a canonical [tʃʰi], they were instructed to give it a goodness score of 1. If, however, the token did not sound like a [tʃʰi] at all, they were instructed to give it a goodness score of 7. Based on our hypothesis, we predicted that the [kʰi] token that was filtered with a high order and wide bandwidth would have better goodness scores than plain [kʰi] (see Figure 4.8).

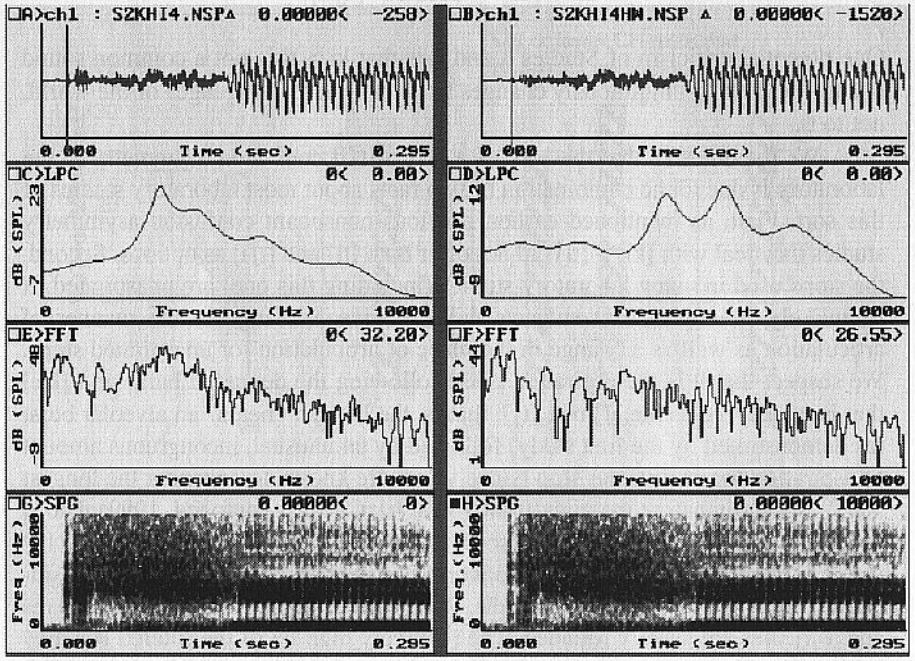

Figure 4.8. Waveform, LPC at burst, FFT at burst, and spectrogram of [k^{h}i] and KHIHW ([k^{h}i] band-reject filtered with a high order, 35, and wide bandwidth, 2 kHz.) Note the characteristic mid-frequency spectral peak of the velar burst in the left column. The spectral peak in the right column has been attenuated by filtering. (Spectral scales have not been normalized.)

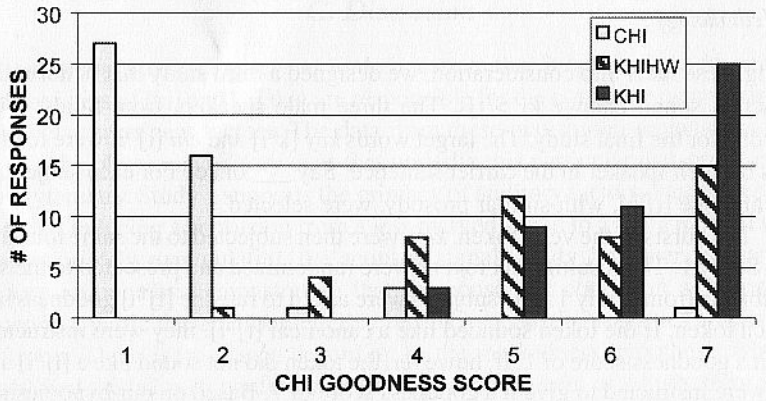

Figure 4.9. [tʃ^{h}i] goodness scores for [tʃ^{h}i] (CHI), [k^{h}i] (KHI), and filtered [k^{h}i] (KHIHW). '1' is the best [tʃ^{h}i] goodness score, and '7' is the worst.

2. Results

As expected, the vast majority of [tʃ^{h}i] tokens received a goodness score of 1 or 2, and most of the natural [k^{h}i] tokens received scores of 6 or 7. As in the first study, only the higher order, wider bandwidth filter resulted in a significantly different goodness score from the unfiltered [k^{h}i] token. These filtered [k^{h}i] tokens, as expected, received better goodness scores than the natural [k^{h}i] tokens. These results are summarized in Figure 4.9 and Table 4.10.

TABLE 4.10

Mean [tʃ^{h}i] Goodness Scores and Standard Deviation for [tʃ^{h}i] (CHI), [k^{h}i] (KHI), and Filtered [k^{h}i] (KHILN, KHILW, KHIHN, KHIHW)

Stimulus	Mean	Standard deviation
CHI	1.6875	1.132804
KHI	6.2083	0.966642
KHILN	6.3125	0.926128
KHILW	6.0833	0.985714
KHIHN	6.1666	0.974861
KHIHW	5.3958	1.41029

'1' is the best score, and '7' is the worst. Of the filtered tokens, only KHIHW displayed a significantly better score.

B. Study 3B

1. Methodology

According to the hypothesis adopted here, that it is the aspiration noise of [tʃ^{h}i] in combination with the mistaken alveolar perception of filtered [k^{h}i] that is responsible for the [ki] > [tʃi] confusion, we would expect that the confusion would be reduced in the case of the voiced counterpart [gi] since there is no aspiration after the burst. To empirically verify this prediction, we replicated Study 3a with voiced tokens.

The same three male speakers from Study 1 were recorded for the final study. The target words ghee [gi] and gee [dʒi] were read five times by each speaker in the carrier sentence 'Say ___ once.' For each subject, one [gi] and one [dʒi], with similar prosody, were selected.

The bursts of the velar token [gi] were then subjected to the same four filters from Study 1. The resulting six tokens were randomized and presented to the same 16 subjects from Study 1. The subjects were asked to rate the [dʒi] goodness score of each token. If the token sounded like a canonical [dʒi], they were instructed to give it a goodness score of 1. If, however, the token did not sound like a [dʒi] at all, they were instructed to give it a goodness score of 7. Based on our hypothesis, we predicted that none of the filtered [gi] tokens would have better goodness scores than plain [gi] (see Figure 4.10).

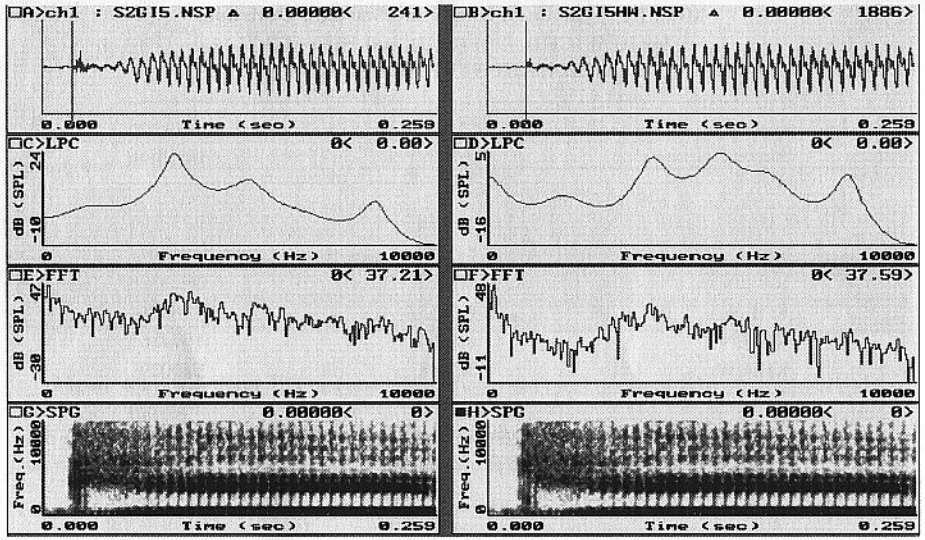

Figure 4.10. Waveform, LPC at burst, FFT at burst, and spectrogram of [gi] and GIHW ([gi] band-reject filtered with a high order, 35, and wide bandwidth, 2 kHz.) Note the characteristic mid-frequency spectral peak of the velar burst in the left column. The spectral peak in the right column has been attenuated by filtering. (Note that spectral scales have not been normalized.)

2. Results

As expected, most of the [dʒi] tokens received a goodness score of 1 or 2, and most of the [gi] tokens received a goodness score of 6 or 7. More importantly, none of the filtered [gi] files, including the one that was filtered with a high order and a wide bandwidth (GIHW), had significantly better scores than natural [gi]. The results are summarized in Figure 4.11 and Table 4.11.

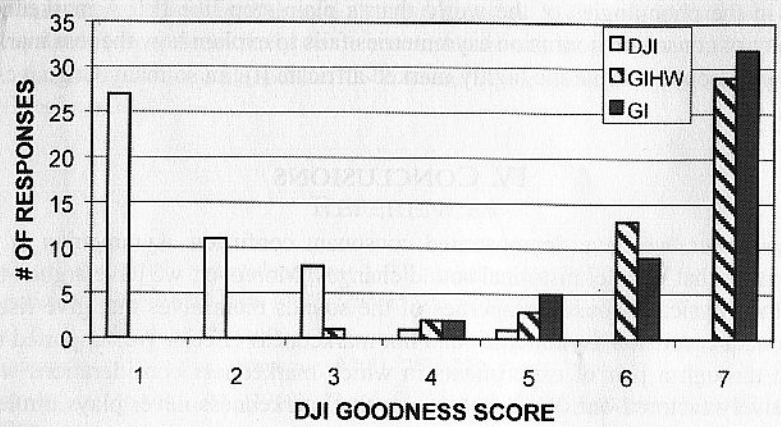

Figure 4.11. [dʒi] goodness scores for [dʒi] (DGI), [gi] (GI), and filtered [gi] (GIHW). '1' is the best [dʒi] goodness score, and '7' is the worst.

TABLE 4.11

Mean [dʒi] Goodness Scores and Standard Deviation for [dʒi] (DJI), [gi] (GI), and Filtered [gi] (GIHN, GIHW, GILN, GILW)

Stimulus	Mean	Standard deviation
DJI	1.7083	0.966642211
GI	6.4792	0.850271129
GIHN	6.3333	1.05856855
GIHW	6.3958	0.939433585
GILN	6.4375	0.872907833
GILW	6.375	1.023656359

'1' is the best score and '7' is the worst.

C. Discussion

Insofar as we were able to influence the direction of consonant confusion by targeted modification of some of the acoustic cues of stops, we believe we have demonstrated that it is acoustic–auditory factors that underlie such perceptual errors. But this study also undercuts claims that markedness plays a major role in these confusions. As far as markedness is concerned, affricates such as [tʃ] are

rarer in the phonologies of the world than a plain stop like [k]. A markedness account of consonant confusion asymmetries fails to explain how the less marked velar stop could became the highly marked affricate [tʃ] in so many languages.

IV. CONCLUSIONS

In this paper, we have demonstrated consonant confusion asymmetries in the laboratory that parallel historical sound changes. Moreover, we have argued that it is the physical, acoustic properties of the sounds themselves that give rise to consonant confusion asymmetries and not markedness effects. We supported this claim through a pair of experiments in which markedness considerations were effectively factored out. We do not argue that markedness never plays a role in consonant confusion asymmetries, but rather that they are secondary to the acoustic-auditory factors. Perhaps, as we hypothesized earlier, markedness effects arise due to the physical acoustic properties of sounds. Finally, we showed that laboratory-based consonant confusion asymmetries do in fact lend insights to actual historical sound changes.

There remain some issues that warrant further scrutiny. First, an investigation of the role of the number of velar bursts, VOT, and burst duration in consonant confusion asymmetries and their relation to velar spectral characteristics may shed more light on the [ki] > [ti] consonant confusion asymmetry. Research currently underway focuses on these issues as well as exploring the relative importance of such cues (in addition to the relative amplitude of bursts versus vowel and formant transition information) in the detection and misperception of stop place in various vocalic environments.

ACKNOWLEDGMENTS

This study was funded by a National Science Foundation grant (#98-17243); Principal Investigator: John J. Ohala. The authors would like to acknowledge Sue-Wen Chiao and Julie A. Lewis for their research assistance and the editors of this volume for helpful comments.

NOTES

1. The parallelism in lab confusion studies and sound change involves the phonetic character of the sounds involved. For example, regarding the Czech palatalized labials, they are phonetically palatalized due to the palatalization being a distinctive secondary articulation whereas in the lab studies involving English the stop in [pi] is palatalized non-distinctively due to coarticulation with the following palatal vowel.
2. In this paper, we use the notation X > Y, where "X can be confused as Y, but Y is rarely confused as X."

REFERENCES

Chang, S. C. (1999). Vowel-dependent VOT variation. *Proceedings of the XIVth International Congress of Phonetic Sciences, San Francisco, 2*, 1021.

Cho, T., & Ladefoged, P. (1999). Variation and universals in VOT: Evidence from 18 languages. *Journal of Phonetics, 27*(2), 207–229.

Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. New York: Harper & Row.

Delogu, C., Paoloni, A., Ridolfi, P., & Vagges, K. (1995). Intelligibility of speech produced by text-to-speech systems in good and telphonic conditions. *Acta Acoustica, 3*, 89–96.

Gilmore, G. C., Hersh, H., Caramazza, A., & Griffin, J. (1979). On the prediction of confusion matrices from similarity judgments. *Perception & Psychophysics, 25*(5), 425–431.

Guion, S. G. (1998). The role of perception in the sound change of velar palatalization. *Phonetica 55*, 18–52.

Lyublinskaya, V. V. (1966). Recognition of articulation cues in stop consonants in transition from vowel to consonant. *Soviet Physics–Acoustics, 12*(2), 185–192.

Miller, G. A., & Nicely, P. E. (1955). An analysis of perceptual confusions among some English consonants. *Journal of the Acoustical Society of America, 27*(2), 338–352.

Plauché, M., Delogu, C., & Ohala, J. (1997). Asymmetries in consonant confusion. *Proceedings of Eurospeech '97: Fifth European Conference on Speech Communication and Technology, 4*, 2187–2190.

Stevens, K. N., & Blumstein, S. E. (1978). Invariant cues for place of articulation in stop consonants. *Journal of the Acoustical Society of America, 104*(1), 1358–1368.

Trubetzkoy, N. S. (1939). *Grundzuge der Phonologie*. Prague.

Wang, M. D., & Bilger, R. C. (1973). Consonant confusions in noise: A study of perceptual features. *Journal of the Acoustical Society of America, 54*(5), 1248–1266.

Winitz, H., Scheib, M. E., & Reeds, J. A. (1972). Identification of stops and vowels for the burst portion of /p,t,k/ isolated from conversation speech. *Journal of the Acoustical Society of America, 51*(4), 1309–1317.