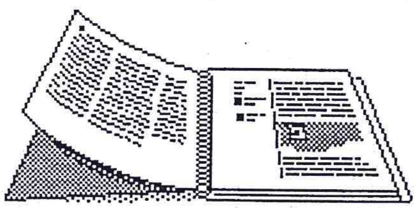
Zentrum für Allgemeine Sprachwissenschaft,
Sprachtypologie und Universalienforschung

ZAS Papers in Linguistics

Volume 28
October 2002



Edited by

T. A. Hall
Bernd Pompino-Marschall
Marzena Rochoń

ISSN 1435-9588

Papers on Phonetics and Phonology:
The Articulation, Acoustics and Perception of Consonants

Edited by

T. A. Hall
University of Leipzig
Bernd Pompino-Marschall
Humboldt-Universität zu Berlin
Marzena Rochoń
Zentrum für Allgemeine Sprachwissenschaft, Berlin

ZAS Papers in Linguistics 28, 2002

Table of Contents
Volume 28

Acoustic Cues for the Korean Stop Contrast - Dialectal Variation

Hansook Choi
University of Illinois at Urbana-Champaign

Abstract

In this study, cross-dialectal variation in the use of the acoustic cues of VOT and F0 to mark the laryngeal contrast in Korean stops is examined with Chonnam Korean and Seoul Korean. Prior experimental results (Han & Weitzman, 1970; Hardcastle, 1973; Jun, 1993 &1998; Kim, C., 1965) show that pitch values in the vowel onset following the target stop consonants play a supplementary role to VOT in designating the three contrastive laryngeal categories. F0 contours are determined in part by the intonational system of a language, which raises the question of how the intonational system interacts with phonological contrasts. Intonational difference might be linked to dissimilar patterns in using the complementary acoustic cues of VOT and F0. This hypothesis is tested with 6 Korean speakers, three Seoul Korean and three Chonnam Korean speakers. The results show that Chonnam Korean involves more 3-way VOT and a 2-way distinction in F0 distribution in comparison to Seoul Korean that shows more 3-way F0 distribution and a 2-way VOT distinction. The two acoustic cues are complementary in that one cue is rather faithful in marking 3-way contrast, while the other cue marks the contrast less distinctively. It also seems that these variations are not completely arbitrary, but linked to the phonological characteristics in dialects. Chonnam Korean, in which the initial tonal realization in the accentual phrase is expected to be more salient, tends to minimize the F0 perturbation effect from the preceding consonants by taking more overlaps in F0 distribution. And a 3-way distribution of VOT in Chonnam Korean, as compensation, can be also understood as a durational sensitivity. Without these characteristics, Seoul Korean shows relatively more overlapping distribution in VOT and more 3-way separation in F0 distribution.

1 Introduction

Phonological contrasts involving the laryngeal features of stop consonants are marked through the interaction of diverse acoustic features (Lisker & Abramson, 1964). Contrasts among oral stops involving voicing (e.g., French), aspiration (e.g., Mandarin), or both (e.g., Hindi) are manifest by the presence or absence of phonation during stop closure and variation in the onset of voicing following stop release (Borden et al., 1994). Many factors may interact to affect the acoustic realization of a phonological contrast in laryngeal features, and prosodic factors are a primary source of variation in many languages. Language varieties that have the same phonological contrast but differ in their prosodic systems might be expected to differ in the realization of the acoustic cues that signal the contrast. This paper explores such differences in the acoustic realization of a 3-way laryngeal contrast in two dialects of Korean that differ in their prosodic systems.

Korean stops exhibit a 3-way laryngeal contrast among oral stop consonants: plain (/tal/, 'moon/month'), tense (/t'al/, 'daughter'), and aspirated (/t^{h}al/, 'mask')[1]. All three categories of stops are voiceless and pulmonic egressive, and differ in a number of acoustic features. VOT and the F0 value at the following vowel onset distinguish at least partially

[1] I follow Han's terms for the three kinds of stops in Korean (Han, 1998). Kim, C. (1997) uses occlusive, tense, and aspirate, whereas Kim, W. (1983) divides them into lenis, fortis, and aspirated. On the other hand, Silva (1991) uses the terms of lax, reinforced, and aspirated.

among the three stop categories. Aspirated stops exhibit the highest VOT and F0 values, while tense stops show the lowest VOT values and a relatively higher F0 onset than the plain series. The laryngeal contrast is also evident in differences in wide-band spectrograms of the initial portion of the periodic signal following release of the three kinds of stops. After tense stops, the voice source is relatively undamped, while damping occurs after aspirated stops, and to the greatest degree, after plain stops (Han & Weitzman, 1970). In the temporal dimension, plain stops show a shorter closure duration than tense or aspirated stops.

Among these various acoustic markers of the Korean laryngeal contrast, VOT has been cited as primary. The success of VOT in distinguishing the three categories is not complete, though. Lisker and Abramson (1964) observe that the three divisions in VOT values in Korean are not perfectly separated from one another. Several subsequent studies confirm the finding of overlapping VOT ranges in Korean, and together they shed doubt on the status of VOT as the primary acoustic correlate of the laryngeal contrast (Han & Weitzman, 1970; Hardcastle, 1973; Kim, C., 1965). Results of VOT analysis are not consistent across these studies, as illustrated by comparing Lisker and Abramson's findings with those of Han and Weitzman, as represented in Figures 1 and 2 below.

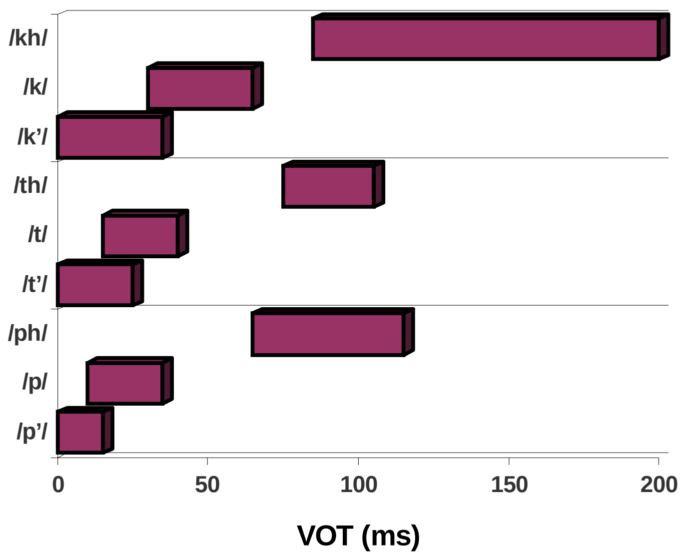
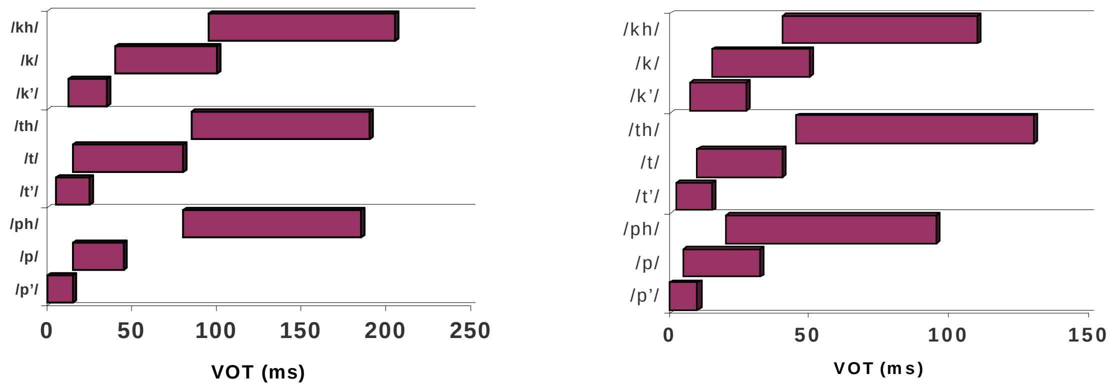


Figure 1. *Lisker and Abramson*'s VOT of Korean stops in initial position (Adapted from Lisker & Abramson, 1964).



(a) Informant 1 (male) (b) Informant 3 (female)

Figure 2. *Han and Weitzman*'s VOT of Korean stops in initial position (Adapted from Han & Weitzman, 1970).

Figure 1 relates the finding from Lisker & Abramson that VOT ranges of plain and tense stops overlap. Figure 2 shows a similar finding from Han & Weitzman (1970). Additionally, Figure 2 displays overlap between the plain and aspirated stops. The tense stop never overlaps with the aspirated stop in VOT in either study. These findings suggest that VOT alone cannot be a sufficient cue for the different laryngeal gestures in the word-initial position.

Overlap in the values of an acoustic cue that marks a phonological contrast might cause confusion for the listener if the cue is the only acoustic feature to mark the contrast. In the case of the laryngeal contrast in Korean, phonation and aspiration gestures also play a role in establishing the phonological contrast. Tense stops have been analyzed as involving a gesture of glottal constriction (marked by [+constricted glottis]), while the aspirated stops involve glottal abduction (marked by [+spread glottis]). Glottal adjustment can be linked to variation in the onset of phonation following the stop release, influencing the F0 value at the onset of the following vowel (Kingston & Diehl, 1994). Evidence for the distinct glottal gestures of the Korean oral stops has been reported by Kim (1965), who observed a slower rate of vocal fold vibration at the beginning of the vowel following the plain stops. This finding is consistent with the general characterizations provided by Hardcastle (1973) and Hombert (1978) that associate aspirated or tense obstruents with a raised F0 of the following vowel onset. Han and Weitzman (1970) also report multiple differences in the onset of voicing after stop release dependent on the laryngeal characteristics of the stop. Among these, F0 is argued to serve as an additional cue to VOT in marking the laryngeal contrast. In general, F0 after plain stops is reported to be lower than the other stop categories (Hardcastle, 1973; Han & Weitzman, 1970; Kim, C., 1965). This general pattern is confirmed further in the subsequent study by Cho (1996), and by my pilot study, where I also observed consistently lower F0 values after tense stops than aspirated stops with the lowest F0 values after plain stops.

Supporting evidence for the role of F0 as a supplementary cue in marking laryngeal contrasts among oral stops is provided by Whalen et al. (1993), who demonstrate that the F0 value at stop release influences perception of the laryngeal feature of the stop even when the primary VOT cue is not ambiguous. Though Whalen et al.'s study concerns the phonological voice contrast, which is not involved in the Korean system, the findings are still significant for Korean because the 3-way Korean contrast is similarly manifest in F0 differences following stop release[2]. Cho (1996) and Kim (2000) also report experimental results which show that F0 perturbation after stop release cues the identification of the laryngeal stop category in Korean.

1.1 Dialectal variation in acoustic characteristics

If VOT and F0 together signal the laryngeal category of stops in Korean, the question arises how these two features interact. The findings of Cho and Ladefoged (1999) show that languages with an analogous laryngeal contrast may nonetheless vary in the manifestation of that contrast through the acoustic feature of VOT. In other words, the phonetic realization of the laryngeal contrast is not predictable from the phonological system of contrast alone. Similar variation may also be expected in the case of distinct varieties of a single language. Moreover, variation in the expression of a single cue to a contrast, such as VOT, raises the possibility about correlated patterns of variation in other cues to the same contrast.

In this paper, I investigate the question of whether distinct varieties of Korean exhibit variation in the way the VOT and F0 cues signal the laryngeal contrasts, and in the interaction between the two acoustic features. This investigation is motivated in part by the observation that among the varieties of Korean for which phonological descriptions exist, there are significant differences in prosodic systems that could in principle lead to different patterns of realization of the laryngeal contrast for stop consonants. In particular, prosodic features that are expressed in patterns of duration and F0 are predicted to have a possible interaction with the use of VOT and F0 to cue the laryngeal contrast. I report here on differences in the acoustic realization of the 3-way laryngeal contrast in word-initial position in two dialects of Korean that differ in prosodic features expressed through pitch and duration patterns. I show that the variation in the acoustic realization of the laryngeal contrast through VOT and F0 cues can be related to the differences in the prosodic systems of the two dialects.

A recent study by Cho et al. (2001) provides additional confirmation of dialectal variation in Korean for the cues that mark the laryngeal contrast. That study compared the acoustic features of the Seoul and Cheju dialects, examining VOT, burst energy, F0 and energy values (dB) for the first and second harmonics and for the peak harmonic of the second formant, and aerodynamic characteristics of oral pressure and flow. Of interest here, Cho et al. report different patterns for VOT and F0 cues: the Cheju dialect exhibits more overlap in VOT ranges for the three types of consonants than does the Seoul dialect, with a compensatory 3-way division in F0 values in Cheju. On the other hand, the Seoul dialect exhibited less overlap in VOT values and only a 2-way distinction in F0 values (plain vs. aspirated and tense). Cho et al. do not elaborate on the possible basis for this dialectal difference, and therefore it is hard to determine if the discrepant patterns of VOT and F0 derive from or are related to any other divergent properties of the two dialects.

[2] Kingston & Diehl (1994) maintain that the similar effects on F0 following stop release between Korean and English support a parallel phonological analysis involving the feature [voice]. Specifically, they argue that the contrast between unaspirated and aspirated Korean stops should be analyzed in terms of the features [+voice] and [-voice], respectively. In support of this argument Kingston & Diehl observe that the Korean unaspirated stops are realized with phonetic voicing in intervocalic position, just as English [+voice] stops are. Under this view, the acoustic property of F0 depression marks the phonological category [+voice] in Korean as it does in English, German, Swedish, and Danish. However, this interpretation of voicing on the basis of F0 depression is made more complicated in light of a more complete picture of the F0 cue to the laryngeal category in Korean. As reported below, for at least some speakers F0 values are distinct for each of the three contrastive laryngeal stop categories in Korean.

1.2 Chonnam Korean vs. Seoul Korean

The Korean dialects compared in the present study are the Seoul and Chonnam dialects, whose prosodic features have been analyzed and described by Jun (1993, 1998). The dialects differ in their intonational systems and in the contrastive status of vowel length. The Chonnam dialect has a phonemic vowel length contrast and a consistent, salient phrase-initial rising intonational contour. The Seoul dialect has no vowel length contrast and a consistent, salient rising intonational contour typically realized phrase-finally. Based on the observed phonological differences between these dialects, there are two hypotheses for this study. First, in Chonnam Korean, the salient phrase-initial pitch contour may restrict the role of F0 in cueing the laryngeal contrast in initial position. Second, the presence of a phonological contrast in vowel length may render Chonnam speakers more sensitive to the differences in the durational properties of vowels in the context of different preceding consonants, supporting a greater role for VOT in signaling the laryngeal contrast. In comparison, Seoul Korean is hypothesized to allow a greater role for F0 as a cue to the laryngeal contrast in phrase-initial position, since there is no interference with a demarcative intonational pattern in that position. On the other hand, Seoul is predicted to have a lesser role for VOT in marking the laryngeal contrast, since the phonological system does not exploit duration for a contrastive feature and thus listeners may not be as well-attuned to durational features associated with vowels. The experiment described in the following section examines the VOT and F0 features of each laryngeal type of consonant in Chonnam and Seoul dialects, and examines the role each feature plays in distinguishing among the laryngeal categories.

2 Methods

2.1 Subjects

Subjects were 6 Koreans, three Seoul Korean speakers and three Chonnam Korean speakers. The subjects were born in the given dialectal area under the parents who are speakers of the same dialect and had stayed in the dialectal region until they moved to the United States. In each dialectal group, two of the subjects were male and one was female. They were similar both in their age, all being in their late twenties, and in the period of time they had been in the States, all less than 2 years. They were all graduate students and reported no speaking impairment. None of them had had any phonetic training and knowledge for this kind of experiments.

2.2 Stimuli

The stimuli consisted of nonsense words with a CVCV sequence, in which the initial C was the one of the target consonants, /p, t, k, p’, t’, k’, p^h, t^h, k^h/. The medial C was /p/. The vowel /a/ was used for V position because it is one of the edge vowels and does not involve phonological phenomenon as Umlaut for /i/ in Korean. To reduce variation from the vowel context, an identical vowel was used for all stimuli. Disyllabic words were chosen because they sound like more natural and common words in Korean. Within the set, the words were different only in the target C as in (1).

(1)	Velar:	/kaba/[3]	/k'aba/	/k^haba/
	Alveolar:	/taba/	/t'aba/	/t^haba/
	Labial:	/paba/	/p'aba/	/p^haba/

All the target words were embedded in the frame sentence, /nanɯn _______ hɛyo/[4] 'I say _________'.

2.3 Procedure

Each subject repeated the nine stimuli 10 times for a total of 90 stimuli per subject. Before the recording procedure, done in a sound-attenuated booth in the phonetics laboratory in University of Illinois at Urbana-Champaign, there was a short instruction session in Korean. For familiarization for each word, subjects rehearsed the word list before recording. Target words were provided visually in a random order, and each repetition was done with a small pause between each word. All repetitions were recorded with a Tascam DAT recorder.

2.4 Measurements

The recorded words were transferred to a PC and analyzed with the Praat program (Boersma & Weenink). VOTs were measured from the stop release to the onset of the second formant in the following vowel[5]. Fundamental frequency value was taken from the onset at the phonation phase after the stop release with the pitch analysis of Praat program, using autocorrelation function. The onset value was taken as the mean of five initial values, which were measured every 0.01 second from the beginning of the second formant.

3 Results

3.1 Results of ANOVA analyses

The results are analyzed for each speaker individually. Because of possible differences caused by gender, male and female subjects were compared within the own gender group.[6]

The results of an ANOVA analysis for each individual showed that the factor of laryngeal contrast had a significant effect for all the subjects ($F(2,88) = 460.891$, $p < .0001$ for Chonnam Male 1; $F(2, 88) = 493.793$, $p < .0001$ for Chonnam Male 2; $F(2,88) = 228.799$, $p <. 0001$ for Seoul Male 1; $F(2,88) = 547.726$, $p <. 0001$ for Seoul Male 2). Tukey and Scheffe post hoc comparison showed that the three laryngeal categories are significantly distinctive from one another for the two Chonnam male speakers and for Seoul male speaker 1, with the highest value for the aspirated, intermediate for the plain, and the lowest value for the tense. Seoul male speaker 2 did not show a clear separation between plain and aspirated stops.

---

[3] Korean voiceless plain stops become voiced without aspiration in intervocalic position, and the phonetic aspects were followed to describe the intervocalic stop.
[4] The phonetic terms for vowels follow Lee's notation in the *Handbook of the International Phonetic Association* (1999).
[5] This is the same measurement technique used in Cho et al. (2001). For the intervocalic plain stops, some of which showed voicing during their closure, the stop release was clearly detected, and VOTs were measured from that point.
[6] Smith (1978), Swartz (1992), and Ryalls et al. (1997) discuss the gender effect on VOT.

Another significant effect was shown by the place-of-articulation for all the subjects (F(2,88)=96.948, p<.0001 for Chonnam Male 1; F(2, 88) = 6.206, p = .003 for Chonnam Male 2; F(2,88) = 15.227, p < .0001 for Seoul Male 1; F(2,88) = 23.771 , p < .0001 for Seoul Male 2). All speakers showed significantly higher VOT values for velar stops than the other stops.

The results of the analysis of variance does not clearly show a dialectal difference. The second Seoul male speaker shows different patterns in VOT from the other three speakers. On the other hand, examination of the distribution of VOT values shown in Figure 3 reveals a distinctive pattern for the two dialects. In Seoul Korean, the VOT range for plain stops overlaps substantially with the range for the aspirated stops. The values for the three stop categories in the Chonnam dialect are more separated in their ranges.

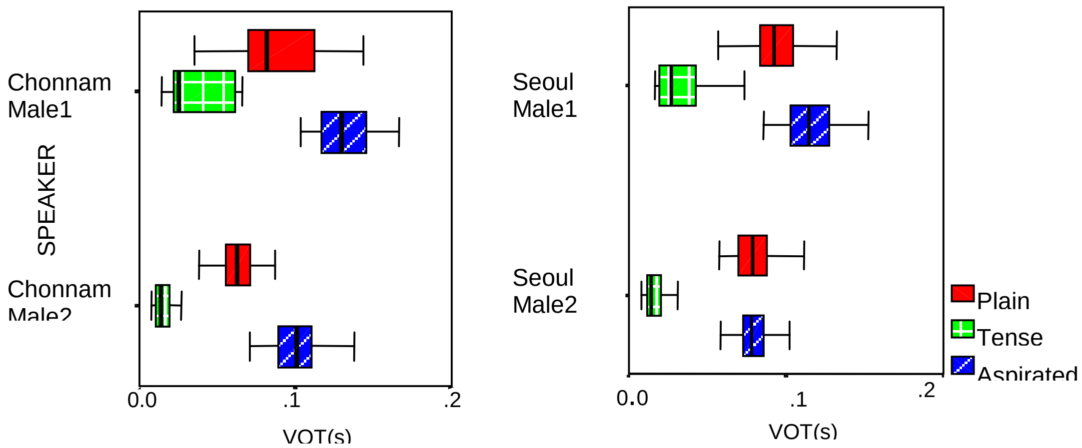


Figure 3. VOT distribution for Chonnam Korean and Seoul Korean.

Figure 3 reveals overlapping VOT values of the three laryngeal categories for all four male speakers. The two dialects are, however, different in the distribution of the interquartile ranges marked with shaded boxes. The three interquartile boxes in the Chonnam dialect are well-separated. Comparatively, the interquartile ranges for plain and aspirated stops in Seoul Korean show a substantial overlap, while the ranges for tense stops are more separated from the other two categories. In other words, compared with Seoul Korean, Chonnam Korean shows more of a 3-way distinction in the range of laryngeal categories of stops, whereas Seoul Korean shows only a 2-way distinction among the three categories.

One-way ANOVA was done for all the subjects separately. The results show a significant effect from the laryngeal contrast for all the speakers (F(2,88) = 135.332, p < .0001 for Chonnam Male 1; F(2, 88) = 540.468, p = .003 for Chonnam Male 2; F(2,88) = 105.384, p < .0001 for Seoul Male 1; F(2,88) = 294.758 , p < .0001 for Seoul Male 2), while there is no effect from place of articulation. The results of a Tukey and Scheffe post hoc comparison indicate that the F0 values for two Seoul male speakers and Chonnam male speaker 2 have 3 divisions with the highest F0 value after aspirated stops, the intermediate F0 values after tense stops, and the lowest one after plain stops. In contrast, the homogeneous subsets for Chonnam male speaker 1 were divided into two parts. For this speaker, tense stops and aspirated stops were not clearly separated. The distribution from the box plots reveals a dialectal difference, specifically, in the overlap between tense and aspirated categories, though there is also a fair amount of individual variation.

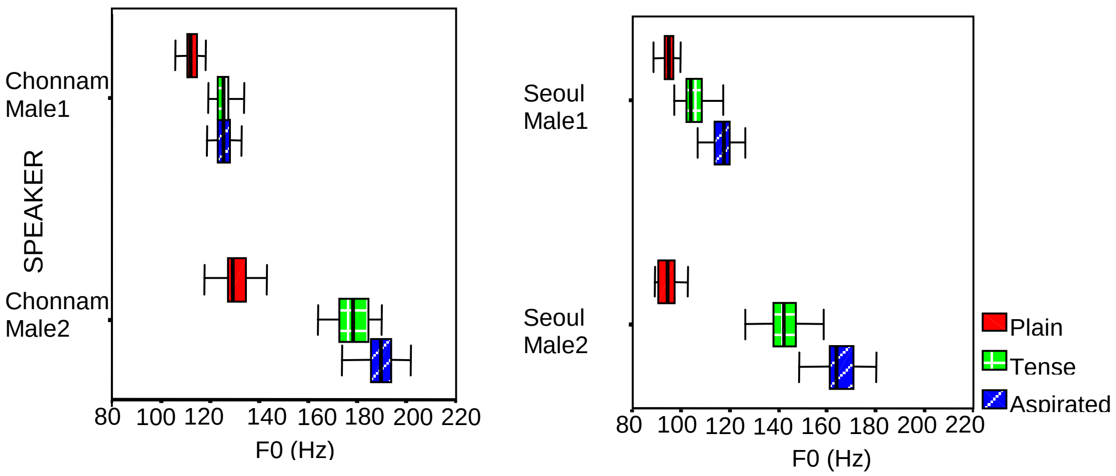


Figure 4. F0 distribution for Chonnam Korean and Seoul Korean.

The Chonnam speakers exhibit a 2-way distinction in F0, with plain stops separated from tense and aspirated stops which overlap. Comparatively, the Seoul Korean speakers show a more evenly spread 3-way distribution of F0 values for the three laryngeal categories.

Since VOT patterns can differ for male and female speech, the female speakers' data are presented separately here. The results of the ANOVA analysis for each individual did not show any difference between the two female speakers. Both female speakers showed a significant effect of laryngeal contrast in F0 and VOT. It is interesting that both female speakers did not show any significant effect from place of articulation in F0 and VOT, though all the male speakers had higher VOT values for velar stops. Tukey and Scheffe post hoc comparison show that there are two homogeneous subsets of values – one for tense stops and the other for plain and aspirated stops – in VOT whereas the distinction for F0 is 3-way for both females. The results are consistent for the two female subjects. But the results of a 2-way ANOVA ([laryngeal contrast] by [speaker]) analysis showed that the interaction between two factors was significant for VOT ($F=93.531$, $P<.0001$) and for F0 ($F=6.842$, $P=.001$). This interaction indicates differences in the distribution of VOT and F0 for the two female speakers.

3.2 Overlapping pattern in the core distribution

For a more precise comparison of the overlapping patterns in the distribution, the percentage of overlap was calculated. The percentage of overlap was considered within the core distribution, because the presence of outliers in a few extreme examples distorts the range. The ranges of the interquartile distribution were taken as the core distribution, which was calculated by putting the mean of the 95% confidence interval as a center value. The entire range of the core distribution had an equal value to that of the interquartile range. In other words, the upper bound of the core distribution was calculated by adding half the value of interquartile range to the mean of the 95% confidence interval, while the lower bound was determined by deducting half the value of interquartile range from the mean of 95% confidence interval. To confirm the normality of distribution, the Kolmogorov-Smirnov test showed that none of the actual distributions were significantly different from the hypothesized normal distribution at the level of 0.01. With the core distribution, the overlapping percentage was calculated in the following way. First, the number of all tokens within the core distribution range was counted for each category, and then, the number of tokens within the overlapping range, which was established between tense stops and aspirated stops for F0 and between plain stops and aspirated stops for VOT for all the speakers, was also counted. The percentage was, therefore, the number of tokens within the overlapping range over the number of tokens within the core distribution multiplied by 100. Now, the dialectal tendency is clear with following bar charts.

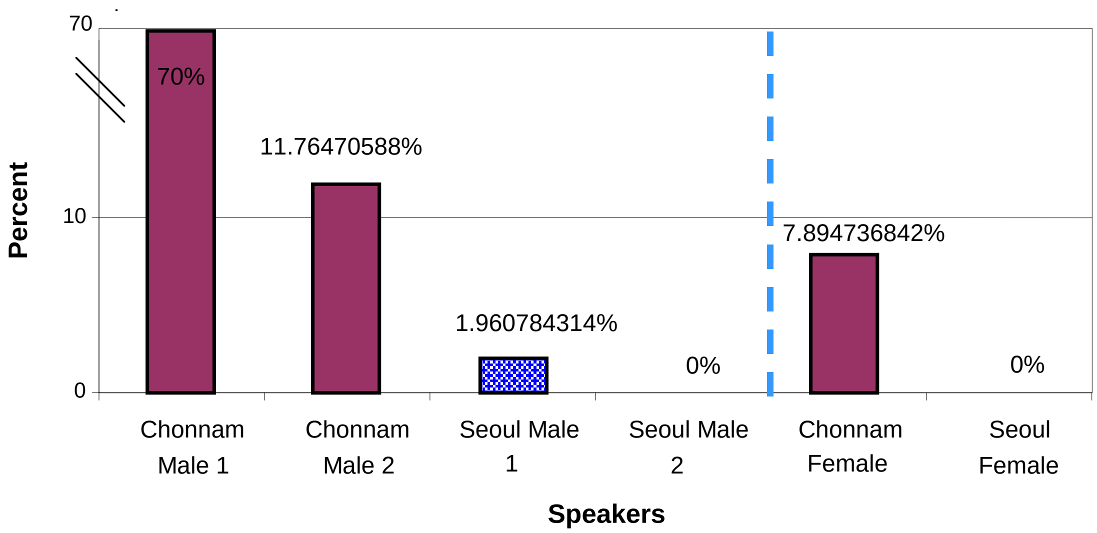


Figure 5. Percentage of overlap in F0 distribution.

In Figure 5, the three bars fully filled with plain color are for Chonnam Korean speakers, and the patterned one is for Seoul Korean speaker. The bar chart shows that there is a noticeable percentage of overlap for all the Chonnam speakers, while only one Seoul male speaker involves minor overlap for F0 values. There is no overlap for the two Seoul Korean speakers. This pattern goes in a reverse way in the following bar chart for VOT values.

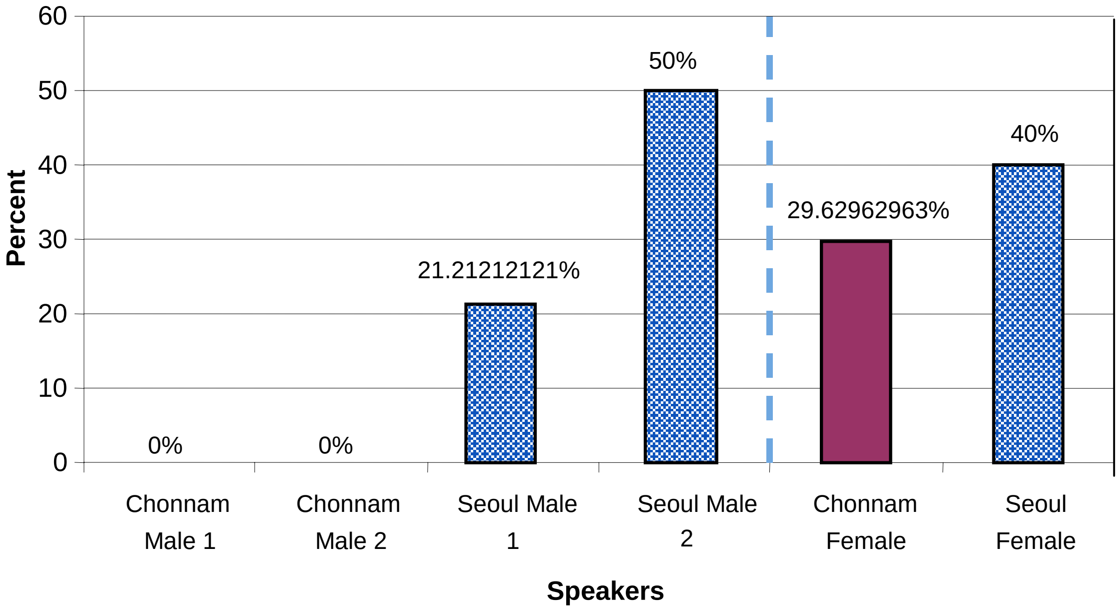


Figure 6. Percentage of overlap in VOT distribution.

Figure 6 shows three bars for substantial percent of overlap in VOT for the Seoul Korean speakers. In contrast, the two Chonnam male speakers do not show any overlap among three laryngeal categories. The Chonnam female speaker still shows the overlapping pattern to some degree but less than the Seoul female speaker's percentage of overlaps. It can be considered as a gender difference in VOT distribution, in that the female speakers show more overlap than the male speakers. Actually, the same pattern that shows more overlap in VOT values can be detected from the floating bars in Figure 2, which is adapted from Han & Weitzman (1970). More research is needed for the topic of gender variation.

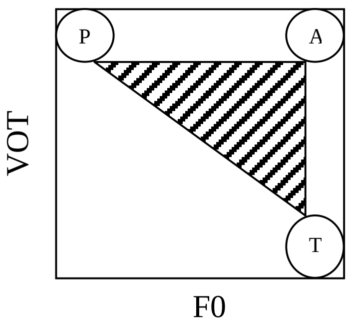
It is, however, noteworthy that the acoustic characteristics of VOT and F0 in the two dialects are complementary, in that Seoul Korean exhibits more overlapping VOT ranges for the three laryngeal categories of stops, which is compensated for by the 3-way distinction in F0. In contrast, Chonnam Korean exhibits 3 separate VOT ranges that mark the three stop categories, and, conversely, there is more overlap in F0 distribution. This complementary patterning in the two acoustic cues is similar to the result reported by Cho et al. (2001) for the Cheju and Seoul dialects.

4 Discussion

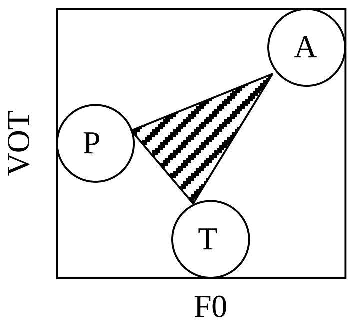
The experimental results reported here show that the same phonological contrast is marked differently in two dialects that differ also in their intonation systems and sensitivity to vowel weight. In Seoul and Chonnam Korean, the acoustic cues of VOT and F0 pattern in a variable but complementary way to mark three stop categories. In summary, there is clear overlap in VOT for all the subjects, and the F0 values show a 3-way distinction for many speakers. As for dialectal tendencies, Seoul Korean shows a 3-way distinction in F0 with more overlapping distribution in VOT values, while Chonnam Korean shows less overlapping VOT distribution for the three stop categories and more overlapping in F0 distribution. The results support the initial hypotheses that the roles of VOT and F0 in signaling the laryngeal contrast will differ according to the interaction with the intonation system and the system of phonological weight contrast.

The patterns of dialectal variation are consistent, but the differences are not categorical. One reason may be the position of the stimuli. They are placed in AP-initial position in both dialects, which conditions some degree of paradigmatic strengthening in Korean (Cho & Jun, 2000; Cho & Keating, 1999; Jun, 1993). The strengthening from the same prosodic boundary may work very similarly in both dialects. The salience of the Chonnam phrase-initial intonation pattern may not observe F0 cues from the consonant, if those cues are strengthened independently in domain-initial position.

It is also noteworthy that individual variation within each dialect is also great. Without mentioning the different ranges of F0 and VOT values, the patterns of overlap or the spread of each category vary to a great degree speaker by speaker. Particularly, the Chonnam female speaker and the first Seoul male speaker are distinguished from the other speakers who display either overlap or no-overlap patterns in F0 and VOT distribution. For these four speakers, a single acoustic cue without overlap can function in a reliable way to demarcate the 3-way laryngeal contrast in stops. On the other hand, the Chonnam female and the first Seoul male speakers show overlap in both cues. None of the acoustic features can be solely reliable to the laryngeal contrast for the two speakers. However, a closer examination reveals that the two speakers are different from each other. The percentage of overlap in the Seoul male speaker is minor for both cues, whereas the Chonnam female speaker shows a great degree of overlap in two cues but a wider range of distribution. Thus, both cues for the first Seoul male speaker are marking the contrast independently, though they are not good enough with overlap and a small separation. On the contrary, the Chonnam female speaker's acoustic cues are dependent on each other marking the 3-way contrast through their interaction. It is very interesting that the 2-dimensional acoustic space defined by F0 and VOT shows a better separation for the Chonnam female speaker, as is illustrated with Figure 7.



(a) 2-way by 2-way (Chonnam Female)



(b) 3-way by 3-way (Seoul Male1)

Figure 7. Simplified representation for 2 dimensional acoustic space with F0 and VOT for consonantal laryngeal contrast in stops. (A: aspirated stops, P: plain stops, T: tense stops)

In Figure 7, the three circles indicate the acoustic values marked with VOT and F0 together, and the slashed triangles inside can be considered as the space between the acoustic representations for the three laryngeal types. Instead of marking a wider range of values for the Chonnam female speaker, the greater separation between the values is depicted with smaller circles in (a). The space of the triangle in (a) is bigger than that in (b), and the distance between plain stops and tense stops in (b) is very small relative to the other separation. The results in Figure 7 also imply that paradigmatic strengthening[7] in acoustic aspects needs to consider the interaction of multiple acoustic cues. That is, in one dimension of VOT or F0, the 2-way separation in (a) may not maximize phonetic distinction of all the three laryngeal categories. But the 2-dimensional space with VOT and F0 in (a) displays a better separation of the three categories than in (b) which seems to enhance the paradigmatic contrast in one-dimensional space of VOT or F0 more than (a).

Perceptual space may be different from acoustic space. If discrepancy in the acoustic patterns corresponds to perceptual patterns, dialectal variation in perception for various acoustic cues is also expected. In future studies, the perceptual patterns for stop identification needs to be tested with speakers from diverse language groups that differs in the acoustic cues that express an analogous phonological contrast.

[7] I understand paradigmatic strengthening as an enhancement of the phonemic distinction of contrastive segments. For example, paradigmatic strengthening marks two contrastive values in a 2-way contrast with a greater separation between the values. In the same way, a 3-way contrast will be marked better with a greater separation among the three contrastive segments under paradigmatic strengthening. Hsu and Jun (1998) show an example of paradigmatic strengthening of VOT in a 3-way contrast

References

Borden, G. J.; Harris, K. S. & Raphael, L. J. 1994. *Speech Science Primer; physiology, acoustic and perception of speech*, 3rd Edition. Williams & Wilkins, USA.

Cho, T. 1996. *Vowel Correlates to Consonant Phonation: An Acoustic-Perceptual Study of Korean Obstruents*. M.A. thesis. University of Texas at Arlington.

Cho, T., Jun, S., & Ladefoged, P. 2001. *Acoustic and aerodynamic correlates to Korean stops and fricatives*. Manuscript.

Cho, T., & Jun, S. 2000 Domain-Initial Strengthening, *CLS* 36, 31-44.

Cho, T. & Keating, P. 1999. Articulatory and acoustic studies of domain-initial strengthening in Korean. *UCLA Working Papers in Phonetics* 97, 100-138.

Cho, T. & P. Ladefoged. 1999. Variations and universals in VOT: evidence from 18 languages. Journal of Phonetics, 27(2), 207-229.

Fowler, C. A. 1996. Listeners do hear sounds, not tongues, *Journal of Acoustic Society of America,* 99(3), 1730-1741.

Han, J.-I. 1998. VOT in the surface distinction of Korean plain and tense stops in initial position: A perception test, *Korean Journal of Speech Sciences, 3,* 109-117.

Han, M. S. & R. S. Weitzman. 1970. Acoustic features of Korean /P, T, K/, /p, t, k/ and /p, t, k/. *Phonetica*, 22, 112-28.

Hardcastle, W. J. 1973. Some observation on the tense-lax distribution in initial stops in Korean, *Journal of Phonetics. 1*, 263-272.

Hillenbrand, J. M. & Houde, R. A. 1996. Role of F0 and Amplitude in the perception of intervocalic glottal stops, *Journal of Speech and Hearing Research, 39,* 1182-1190.

Hsu, C. & Jun, S. A. 1998. Prosodic strengthening in Taiwanese: Syntagmatic or paradigmatic? *UCLA Working Papers in Phonetics* 96, 69-89.

Hwang, H.-Y. 1979. *An Introduction to Korean Phonology,* Sam-wo-chool-pan-sa, Seoul.

Jun, S. A. 1993. *The Phonetics and Phonology of Korean Prosody*, Ph.D. dissertation, Ohio State University.

Jun, S. A. 1998. The Accentual Phrase in the Korean prosodic hierarchy. *Phonology*, 15(2), 189-226.

Kim, C. W. 1997. The structure of phonological units in Han'gul, *The Korean alphabet*, 145-160. University of Hawai'i Press.

Kim, C. W. 1965. On the anatomy of the tensity feature in stop classification (with special reference to Korean stops), *Word 21*, 339-359.

Kim, M. R. 2000. *Segmental and Tonal Interactions in English and Korean: A Phonetic and Phonological Study*. PhD dissertation. University of Michigan.

Kim, W. J. 1983. Phonological structure of the Korean language, *The Korean language*, 157-170. Pace International Research.

Kingston, J. & Diehl, R.L.. 1994. Phonetic Knowledge, *Language,* Vol.70, No.3, 419-54.

Lee, H. B. 1999. *Handbook of the International Phonetic Association; A Guide to the Use of the International Phonetic Alphabet*, 120-3.

Lisker, L. & Abramson, A. S.. 1964. A cross-language study of Voicing initial stops: Acoustical measurements, *Word, 20,* 384-422.

Moon, S.-J. & Lindblom, B. 1994. Interaction between duration, context, and speaking style in English stressed vowels, *JASA* 96, 40-55.

Ryalls, J.; Zipprer, A. & Baldauff, P. 1997. A Preliminary Investigation of the Effect of Gender and Race on Voice Onset Time, *Journal of Speech, Language, and Hearing Research,* 40, 642-645.

Silva, D. J. 1991. A phonetically based analysis of [voice] and [fortis] in Korean, *Japanese Korean Linguistics, 2,* 164-174. Stanford University.

Smith, B. 1978. Effects of place of articulation and vowel environment on voiced stop consonant production. *Glossa, 12,* 163-175.

Swartz, B. 1992. Gender differences in voice onset time. *Perceptual and Motor Skills,* 75, 983-992.

Whalen, D. H.; Abramson, A. S.; Lisker, L. & Mody, M. 1993. F0 gives voicing information even with unambiguous voice onset time. *JASA* 93(4), 2152-2159.

Retroflexion and Retraction Revised*

Silke Hamann, OTS Utrecht

Abstract

Arguing against Bhat's (1974) claim that retroflexion cannot be correlated with retraction, the present article illustrates that retroflexes are always retracted, though retraction is not claimed to be a sufficient criterion for retroflexion. The cooccurrence of retraction with retroflexion is shown to make two further implications; first, that non-velarized retroflexes do not exist, and second, that secondary palatalization of retroflexes is phonetically impossible. The process of palatalization is shown to trigger a change in the primary place of articulation to non-retroflex. Phonologically, retraction has to be represented by the feature specification [+back] for all retroflex segments.

1 Introduction

Retroflex consonants are usually described as sounds articulated with a bent-backwards tongue tip and a postalveolar place of articulation, e.g. by Catford (1977: 150) or Trask (1996: 308). As illustrated in Hamann (in prep.), a number of retroflex segments deviate from this definition. The retroflex stops and nasals in Hindi, for example, do not show a bending backwards of the tongue tip but these sounds are still considered to be phonetically and phonologically retroflex. The same holds for the voiced and voiceless postalveolar fricatives in Polish and Russian (ibid.), which are articulated without active involvement of the tongue tip in the constriction (though the tip is raised from its resting position) and with a flat tongue body.

The class of retroflexes shows considerable articulatory variation regarding both the articulator (i.e. apical or subapical to laminal), and the place of articulation (i.e. alveolar to palatal). This variation makes it difficult to find common articulatory properties that hold for all instances of retroflexion.

In his article 'Retroflexion and Retraction', Bhat (1974) discusses the property of retraction as a defining criterion for retroflexion. He states that most retroflexes are retracted, and this retraction can explain both the occurrence of retroflexes with back vowels and their incompatability with front vowels. Retroflex segments and back vowels are both assigned the feature value [+back], and front vowels the feature value [–back]. Nevertheless, Bhat argues that retraction is not a criterion that holds for all retroflexes and furthermore not for retroflexes exclusively. As an example of languages with non-retracted retroflex segments, he mentions the vowel system of the Dravidian language Badaga and the consonantal system of Tamil. As evidence for the non-exclusiveness of retraction to retroflexes, Bhat mentions that apical alveolars[1] are also retracted.

The present study agrees with Bhat's second statement that retraction holds not exclusively for retroflexes but that all apicals (and velars) are retracted. In contrast to Bhat it

* The ideas for this article emerged from two talks on the complementary nature of palatalization and retroflexion held at at the33rd Poznań Linguistic Meeting, Poznań, and the Uil-OTS Utrecht, in April and May 2001, respectively. I want to thank the audience for their input. Furthermore I am grateful to T.A. Hall for comments on an earlier version of this study. All errors are my own.

[1] 'Apical alveolars' is used here in general for all apical coronals which are [+anterior], including apical dentals.

is claimed here that all retroflexes are retracted. This claim is formalized in the implication in (1), which says that if a segment is retroflex, then it is retracted (but not vice versa).

(1) retroflex $\rightarrow$ retracted

From the implication in (1) it follows that there are no non-retracted retroflexes, and consequently so-called non-retroflex segments in Tamil and Badaga are claimed to be phonetically non-retroflex. The inherent retraction of retroflexes further implies that retroflexes cannot be palatalized, because the articulatory gestures of retraction and palatalization are incompatible. It will be argued that secondarily palatalized retroflexes either undergo a change in primary place of articulation (and in segment class) or resist palatalization altogether.

The study proceeds as follows. In section 2, the term ‘retraction’ is articulatorily defined and its correlation with retroflexion is illustrated. Section 3 is concerned with supposedly non-retracted retroflexes in Lardil and Badaga. In section 4, the claim that palatalization and retraction are incompatible is elaborated and apparent counterexamples (Kashmiri and Tamil) are discussed. Section 5 discusses the consequences of retroflexes being retracted for a phonological representation. The final section concludes.

2 Definition of ‘retraction’

Bhat (1974) defines retraction as the backing of the tongue body. This backing can be further specified by the place in the vocal tract where the tongue retracts to, namely towards the pharynx or the velum. The former is often referred to as pharyngealization and the latter as velarization. Thus, the term retraction seems to subsume two secondary articulations. The term pharyngealization is usually used to refer to a secondary vowel articulation (e.g. by Ladefoged & Maddieson 1996: 365) where the root of the tongue is drawn back and towards the back wall of the pharynx. Velarization is understood as a secondary articulation where the tongue back is raised towards the velum (e.g. by Laver 1994 or Trask 1996: 374). In some phonetic literature, the term ‘velarization’ is actually used in the more general sense of retraction. Brosnahan & Malmberg (1970: 67), e.g., define velarization as ‘the elevation of the back of the tongue toward the soft palate or rear wall of the pharynx’, which actually covers both velarization and pharyngealization as defined above.

Ladefoged (1971: 208) points out that there is little articulatory difference between velarized and pharyngealized sounds and no language distinguishes between these two. Language-specific evidence for the interchangeable use of pharyngealization and velarization comes from Russian. Russian speech sounds are traditionally described as opposing palatalized and velarized consonants. Bolla (1981), however, describes the latter as ‘pharyngealized’ because he ‘found the movement of the root of the tongue and the postdorsum towards the pharyngeal wall to be more important than that towards the soft palate’ (p. 70).

‘Retraction’ is defined in the present article as a displacement of the tongue dorsum or root towards the pharynx or velum. A schema of these possible displacements is given in 0, based on x-ray tracings of velarized and pharyngealized segments in Laver (1994: 326ff.) and Ladefoged & Maddieson (1996: 365).

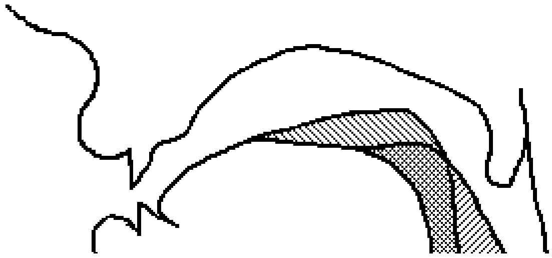


Figure 1: Retraction, indicated as shaded areas, in comparison to neutral tongue position. The upper displacement is a velarization gesture, the lower pharyngealization.

The legitimacy of combining the distinct secondary articulations pharyngealization and velarization as one property is further confirmed by the fact that they are described as resulting in the same acoustic consequence, namely a lowering effect on the third formant (Brosnahan & Malmberg 1970; Stevens 1998). Both velarization and pharyngealization occur with a flattening of the tongue body, which is included here as a characteristic of ‘retraction’.

The property ‘retraction’ is not identical to the feature ‘retracted tongue root’ (opposed to ‘advanced tongue root’, see Halle & Stevens 1969), because this articulatory setting involves a pharyngeal constriction at a lower place in the vocal tract than for pharyngealization (Laver 1994: 411). Furthermore, retraction as defined here is different from McCawley’s (1966) feature ‘retracted articulation’ by which he distinguishes dentals from alveolars and retroflexes from palatals (amongst others), the second item in each pair being [+retracted] (and thus ascribing retroflex a non-retracted status).

The co-occurrence of retroflexion with retraction as predicted in (1) can be explained articulatorily. The tongue, in order to be able to move its tip upwards, stretches and pulls the muscles backwards (Spencer 1984: 30), which results in a backed and flattened or even hollowed tongue dorsum. Further evidence for the correlation of a flat tongue body with retraction and retroflexion can be found in Ladefoged (1971: 208) who explains that the similarity in quality between retroflex stops and velarized or pharyngealized stops “is due to the fact that in all these sounds the front of the tongue is somewhat hollowed” (ibid).

The flattening of the tongue body and its retraction in retroflex segments are also mentioned in Brosnahan & Malmberg (1970: 46) who explain that retroflexion is accomplished by the concaving of the dorsum and blade, i.e. by a flattening of the tongue dorsum. Catford (1977: 157) illustrates that retroflex fricatives have a tongue body that is less convex than in lamino-postalveolar articulations and that they show some velarization. Language-specific descriptions showing a correlation between retroflexion and retraction include Polish and Bulgarian. Hamilton (1980: 21), for example, uses the terms velarization and retroflexion interchangeably for the postalveolar fricative in Polish, and Wood (1996) writes that the Bulgarian retroflexed /r/ involves a pharyngeal tongue body gesture.

Bhat (1974) argues that retraction is not restricted to retroflexes but that it also occurs with apical alveolars. Stevens et al. (1986: 436) argue against this view and claim that a fronted tongue body provides a more favorable posture for an apico-alveolar articulation, i.e. that a non-retroflex apical usually occurs with fronted, non-backed tongue body. This is illustrated with schematic midsagittal sections (ibid. figure 20.4). Below these schemas, the authors say that the tongue-body shape behind to the tongue blade has been only estimated, i.e. no x-ray or other data are given to confirm their claim. X-ray tracings of apicals (such as found in Ladefoged & Maddieson 1996 or Butcher 1992) illustrate that apicals generally show a retracted tongue dorsum.

Retraction of the tongue body (towards the velum) seems to occur distinctively in the articulation of velar consonants (see e.g. x-ray tracing of velar stops and fricatives in Laver

1994). Thus retraction cannot be assumed to occur exclusively with retroflexes. But whereas Bhat discussed retraction as the only defining criterion for retroflexion and refutes this criterion with the argument that it does not distinguish retroflexes from apicals, the present study argues that retraction is only one of several properties of retroflexion. Additional phonetic properties for retroflexes such as apicality, posteriority and sublingual cavity are developed in Hamann (in prep.). Independent of the status of these defining properties, retraction is a necessary property of retroflexe, i.e. it cannot be lacking in a retroflex. In conclusion, retroflex segments are always retracted.

3 Non-retracted retroflexes? The cases of Badaga and Lardil

Using retraction as a defining property of retroflexion poses problems with languages that are said to have a distinctively non-retracted (i.e. non-velarized or non-pharyngealized) retroflex segment.

According to Bhat (1974: 234), the Dravidian language Badaga contrasts plain, retroflex, and retracted vowels. Badaga is the only known language to employ such a contrast. Using retraction and retroflexion contrastively implies that these two properties do not co-occur together in the retroflex segments of this language. Bhat bases his description of Badaga vowels on Emeneau (1939), who describes the three-way contrast however as one of 'non-retroflexed, half-retroflexed, and fully-retroflexed vowels'. Examples for this are given in (2), with Emeneau's transcription.

(2)	Plain vowel		half retroflexed		fully retroflexed	
	kae	'unripe fruit'	*áé*	'tiger's den'	*käë*	'weeds'

Bhat justifies his re-classification of Emeneau's 'half-retroflexed' vowels as 'retracted' by the fact that they are described as having the edges of the tongue tip curved upwards only, but not the whole tongue, and that they show strong retraction (p. 234). As no definition of retroflexed vowels is given which says that these segments have to include a curving of the whole tongue tip backwards, Bhat's classification is not motivated. We assume therefore Emeneau's classification to be correct, which poses no counterevidence for the assumption that the characteristic 'retraction' holds for retroflex vowels, too.

A far more serious counterexample to the claim that retroflexion always co-occurs with retraction seems to be the Australian Aboriginal language Lardil. Lardil is supposed to have phonetically and phonologically non-velarized retroflex consonants (Wilkinson 1988; Hall 1997, 2000). The phonetic evidence for this claim comes from Hall (1997: 49) who gives midsagittal tongue tracings of non-velarized and velarized retroflexes from the languages Lardil and Polish, respectively. The source for the figure of the non-velarized retroflexes in Lardil is said to be Stevens et al. (1986). Stevens et al., however, do not provide any graphic illustrations of the Lardil coronal. The only figure that might have served as basis for Hall's graphic is a schematized retroflex stop in figure 20.4 on page 433, based on Ladefoged & Bhaskararao (1983) and Wierzchowska (1971). These authors describe retroflexes in Hindi, Tamil and Polish. The schematic figure of a retroflex based on their work shows a distinct backing of the tongue body and is explicitly described as 'more backed' than an apical dental or a laminal postalveolar in the following text. In their phonetic description of Lardil sounds at a later point, Stevens et al. (p. 444f.) observe that [+distributed, –anterior] segments, i.e. laminal postalveolars, are [–back] in this language. Hall (1997) presumably misapplied this correlation to [–distributed, –anterior], i.e. retroflex, segments and concluded that Lardil is a language with [–back], i.e. non-velarized, retroflexes.

Phonological evidence for the non-retracted status of the Lardil retroflexes is given in Wilkinson (1988) and Hall (1997, based on the former). Lardil allows only a subset of coronals in coda-position, namely all apart from the lamino-dental [t̪, n̪] and the postalveolar glide [j]. The whole coronal inventory of Lardil is given in table 1 (based on Wilkinson 1988).[2]

Table 1: Coronals in Lardil, boldface ones allowed in Coda position

	lamino-dental	apico-alveolar	lamino-postalveolar	apico-domal (retroflex)
obstruents	t̪	**t**	**t̠**	**ʈ**
nasals	n̪	**n**	**n̠**	**ɳ**
lateral		**l**		**ɭ**
glides			j	**ɻ**
flap		**ɾ**		

To account for the class of [t̪, n̪, j] which cannot occur in coda position, Wilkinson makes use of the feature [back] and assigns the value [+back] only to the lamino-dentals. This yields a feature specification of Lardil coronals as given in table 2.

Table 2: Feature specification of Lardil coronals with [back] according to Wilkinson (1988: 327)

	lamino-dental	apico-alveolar	lamino-postalveolar	apico-domal (retroflex)
[back]	+	–	–	–

In this classification, the lamino-dentals [t̪, n̪] are united under one feature, and thus the phonotactic restriction in Lardil can be described as forbidding [+back] segments in coda-position. This proposal makes it necessary to specify the retroflex segments of Lardil as [–back], which has to be phonetically interpreted as non-retracted. Wilkinson bases this feature specification on Stevens et al. (1986)3, but these authors illustrate that retroflexes have a more backed tongue body than laminal postalveolars (recall discussion above), i.e. they do not confirm Wilkinson's assumption about the non-retraction of Lardil retroflexes. As a further point of criticism it has to be added that Wilkinson's class of [+back] segments does not include the postalveolar glide [j].

As justification for the use of the feature [back], Wilkinson (1986: 328) discusses a process whereby the apical alveolar /t/ becomes a laminal postalveolar [t̪] before the back vowel [u]. The only example given for this is the underlying /yaɻput/ 'snake, bird', which becomes [yaɻput̪uɻ] in the future and [yaɻput̪aɾ] in the marked nonfuture.4 This example is not convincing for two reasons. First, no evidence for the underlying form being an alveolar is given. And second, the spread of the feature [+back] onto the alveolar can be caused by a following back vowel [u] only in the case of the future form, where the morpheme [uɻ] is added. The marked nonfuture morpheme [aɾ] does not contain a back vowel and thus no

[2] The use of the IPA symbols is my own. The term 'laminal postalveolar' is based on Evans' (1995: 727f.) description of Lardil, which is phonetically more detailed than Wilkinson, who describes this class as 'laminal alveolar'. Wilkinson furthermore does not include the retroflex lateral in her inventory.
[3] Wilkinson refers to "Kawasaki, Keyser, and Stevens (1986)" as the source, but the reference list shows that this is meant to be Stevens et al. (1986).
[4] As Wilkinson (1988: 325) explains, nouns in object position agree in tense with the verb.

feature [+back] that could trigger a change, but still this form surfaces with a laminal postalveolar.

An alternative solution to unite the Lardil coronals that do not occur in coda-position is by referring to the features [distributed] and [anterior]. A classification by these features is given in table 3.

Table 3: Feature specification of Lardil coronals with [anterior] and [disributed]

	lamino-dental	apico-alveolar	lamino-postalveolar	apico-domal (retroflex)
[anterior]	+	+	–	–
[distributed]	+	–	+	–

Now, the class of [t̪, n̪, j] can be described as [+anterior, +distributed] stops and nasals, and [+distributed] glides.

In sum, it is phonologically not necessary and phonetically not motivated to describe the class of Lardil retroflexes as [–back]. As no other counterexample with a non-retracted retroflex is known to me, no language with a non-retracted retroflex seems to exist, which affirms the implication made before that retraction always co-occurs with retroflexion.

4 Retraction and palatalization

Besides implying that non-velarized or non-pharyngealized retroflexes do not occur, the retroflex property ‘retraction’ introduced here has a further implication. If retroflex segments are inherently retracted, they should not be compatible with secondary palatalization, because a simultaneous articulation of palatalization and velarization or pharyngealization is articulatory incompatible. An explanation for this incompatability comes from articulation. The palatalization of consonants involves a raising of the tongue dorsum and a lowering of the tongue back, whereas retraction has the opposite articulatory consequences of flattening the tongue dorsum and raising the back. Both gestures cannot co-occur together.

Evidence for the claim that palatalized retroflexes are non-existent is found in Maddieson’s (1984) typological study, which lists no language with a phonemic palatalized retroflex segment. Only two counterexamples could be found in the phonetic and phonological literature, namely Toda (Emeneau 1984; Spajić et al. 1996) and Kashmiri (Bhat 1987), which are both said to have palatalized retroflexes. It is argued in this section that these segments are not retroflex, and that the process of palatalization triggers a change in the retroflex segment from apical to laminal (as proposed already in Hall 2000) and from flat tongue dorsum to bunched tongue dorsum. The second change implies the loss of retraction, thus the resulting segment is ssumed to be non-retroflex.

This section proceeds as follows. In 4.1, traditional definitions of palatalization as mere additional articulations are shown to be inadequate for coronal sounds. Then, to illustrate the change from retroflex to non-retroflex occurring with palatalization, the Russian retroflex fricative and its palatalized counterpart are discussed. In 4.2, the alleged palatalized retroflex segments in Toda and Kashmiri are treated and the status of these segments is analyzed. Alternative descriptions for these supposedly palatalized retroflexes are proposed and it will be concluded that there are no counterexamples for the claim that secondary palatalization of retroflexion does not occur.

4.1 Palatalization as a change in primary articulation

Palatalization in traditional articulatory terms is defined as the superimposition of an [i]-like gesture upon a labial, dental, alveolar or postalveolar consonant (e.g. Ladefoged & Maddieson 1996). This superimposition of a gesture is undoubtedly the case for labials with a secondary palatalization, where the tongue dorsum gesture can take place independently and at the same time as the labial closing gesture. But for primary gestures with the tongue tip, blade or dorsum, the primary and secondary gestures are not independent of each other and therefore are expected to influence each other, which results in a change of the primary place of articulation. Support for this asssumed change can be found in Ladefoged (1971: 207) who points out that “the terms palatalization and palatalized may also be used in a slightly different way from a secondary articulation, namely as describing a process in which the primary articulation is changed so that it becomes more palatal.” Ladefoged & Maddieson (1996: 365) further specify this by stating that for all coronal consonants, secondary palatalization always involves a displacement of the surface of the tongue. This displacement is said to produce a slightly different primary constriction location (ibid.). It is concluded from this, that the traditional description of a secondary palatalization is inaccurate in the case of coronal segments, as this process always involves a change in the primary articulation for coronal sounds.

Articulatory evidence for a change of place in palatalized apical dentals is given in Scatton (1975) for Bulgarian and Čavar & Hamann (2002) for Polish. Hall (2000) argues that apical stops in general either turn into laminal stops when palatalized (in a synchronic or diachronic process), or resist palatalization altogether.

For retroflex segments, it is proposed here that the addition of a palatalization gesture also involves a change in primary articulation. This change results in a non-retroflex segment. Support for this proposal can be found in Ladefoged (1971: 208), who mentions that the secondary articulations of palatalization, velarization and pharyngealization involve different shapes of the tongue that cannot occur simultaneously. As velarization and pharyngealization were defined as realizations of the retroflex criterion ‘retraction’ in section 1, Ladefoged’s remark can be interpreted as an articulatory incompatibility of retroflexion and palatalization.

The incompatibility of gestures and the change in primary place is exemplified with the Russian fricatives in the postalveolar region. Figure 2 is based on x-ray tracings of the Russian retroflex fricative (solid line) and its palatalized counterpart (dashed line) (both based on Bolla 1981: 159). As illustrated in Hamann (2002), the Russian postalveolar fricative is a retroflex which satisfies the property of retraction.[5]

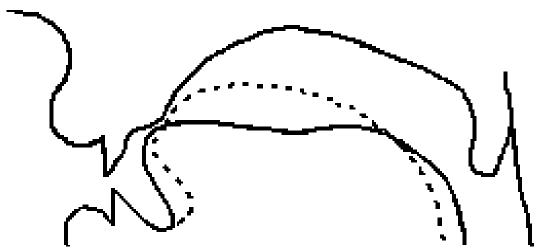


Figure 2: Russian retroflex fricative (solid line) and palatalized postalveolar fricative (dashed line)

Comparing now the palatalized variant with the retroflex one, some major differences can be observed. First of all, the place of articulation changes for the palatalized segment; it moves

[5] Hamann (2002) argues that the Russian voiced and voiceless postalveolar fricatives satisfy three of the four properties for retroflexion introduced there, namely apicality, sublingual cavity and retraction, and is therefore retroflex.

further backwards to the postalveolar region, which gives evidence for the assumed change in primary articulation for retroflex segments. Furthermore, the articulator is now the tongue blade, and the tongue dorsum shape changes to bunched and raised. The resulting segment has a domed tongue, hence no retraction is discernible. This leads to the conclusion that the resulting segment is not retroflex, as it does not fulfill the necessary property of retraction.

As palatalization in general involves the addition of or change towards an [i]-like gesture, and [i] and other front, high vowels are always articulated with a bunched tongue dorsum, this implies that secondary palatalization of retroflexion always results in a change in the property retraction. But as apicality is assumed to always co-occur with retraction, a further change from apical to non-apical is necessary. The process of secondary palatalization of a retroflex segment is therefore assumed to result in a laminal postalveolar. For the fricative [ʂ] this results in the palato-alveolar fricative [ʃ^{j}], cf. (3a). The secondary palatalization of a retroflex stop and nasal is assumed to trigger likewize changes, see (3b) and (3c), respectively.

(3) (a) [ʂ^{j}] = [ʃ^{j}]
    (b) [ʈ^{j}] = [t̻] or [c]
    (c) [ɳ^{j}] = [n̻] or [ɲ]

Hume (1994) assumes that the palatalized postalveolar in Polish is a palatalized retroflex [ʂ^{j}]. At the same time, she specifies this sound as [coronal, -anterior, +distributed, +strident]. The specification as [+distributed] indicates that this sound is laminal, i.e. articulated with a long constriction, and not retroflex. Therefore Hume's description implies that the process of palatalization in Polish actually changes a retroflex fricative into a laminal postalveolar, as stated in (3).

Besides the categorical change described above, another possible outcome of secondary palatalization of retroflexes is to resist palatalization at all, as pointed out by Hall (2000). He gives an example from Scots Gaelic, where nouns usually undergo palatalization in the genitive singular, e.g. [k^{h}at^{h}] 'cat' (nom.sg.) surfaces as [k^{h}at^{hj}] 'cat' (gen.sg.). Nouns with retroflex consonants, however, remain unpalatalized, e.g. [paːʈ] 'a poet' (both nom. and gen. sg.) (Borgstrøm 1940: 76). A resistance to palatalization is otherwise only reported for apical alveolars or dentals (e.g. Hall 2000), which were described as also being inherently retracted in section 1. The property retraction can be made responsible for blocking palatalization, for the same articulatory reason that causes this property to change into non-retraction in secondary palatalized retroflexes, namely articulatory incompatability. As all apicals are retracted, this explains why apicals in general change category when palatalized or resist palatalization altogether.

In sum, it was shown that retraction is incompatible with palatalization, which results in two possible outputs for retroflex palatalization, either a corresponding palatalized laminal, or a plain retroflex without palatalization.

4.2 Apparent counterexamples: Toda and Kashmiri

According to Emeneau (1984), and Spajić et al. (1996), the Dravidian language Toda employs palatalized versions of all its three rhotics, including the retroflex trill /ɽ/. Minimal pairs such as [oɽ] 'to cook' vs. [oɽ^{j}] 'foot', or [toɽ] 'thigh' vs. [toɽ^{j}] 'pole used at funeral' illustrate the alleged contrast between plain and palatalized retroflex rhotic. Interestingly, Spajić et al. (1996) could elicit retroflex rhotics and their palatalized counterparts only from some of their subjects; the three speakers of the Kas mund (tribal location). The three speakers of the

Melgas mund did not produce any of these forms.[6] Though presenting a detailed phonetic study of these rhotics, Spajić et al. (1996) do not include any palatographic or linguographic measurements of the palatalized trill /ɽʲ/, from which the exact articulation and the correlation of the gesture of retroflexion and that of palatalization can be judged. In order to attest the claim made here that the two gestures do not occur simultaneously, articulatory measurements on the Toda palatalized rhotics have to be conducted in the future.

Palatalized retroflex segments are said to also occur in the Indo-Aryan language Kashmiri (Bhat 1987: 43ff.). Kashmiri has the phonemes /ʈʲ, ʈʰʲ, ɖʲ/. Wali & Koul (1997: 297) illustrate these phonemes with the minimal pairs given in (4).

(4)	/həʈʲ/	'throats'	vs.	/həʈ/	'piece of wood'
	/ʈəːʈʰʲ/	'dear ones (m.pl.)'	vs.	/ʈəːʈʰ/	'dear one (f.)'
	/bəɖʲ/	'big ones (m.pl.)'	vs.	/bəɖ/	'big (f.sg.)'

Maddieson's (1984) phoneme inventory of Kashmiri (based on Kelkar & Trisal 1964) does not include palatalized retroflex phonemes; the only retroflexes given there are the plain plosives /ʈ , ʈʰ, ɖ/. Morgenstierne (1941) also does not mention palatalized retroflexes in Kashmiri. The cause for this discrepancy in the description of the Kashmiri phoneme inventory are probably the so-called mātrā vowels. Mātrā vowels are extremely short (Maddieson 1984: 271 terms them 'overshort') or 'whispered' vowels (Masica 1991: 121), which cause changes in the quality of the preceding vowels and consonants. The –i-mātrā is said to leave a palatalizing effect on the preceding consonant. The assumption of such a short /i/ vowel makes the postulation of separate palatalized consonants redundant. Thus, in some descriptions of Kashmiri, e.g. Grierson (1911) and Morgenstierne (1941), the use of a retroflex segment with a following –i-mātrā stands for what is described as palatalized retroflex by others, e.g. Bhat (1987) and Wali & Koul (1997). Bailey (1937) even uses a retroflex plus a full-length vowel /i/ in his transcriptions of these segments in Kashmiri. In (5), a comparison of the different transcriptions is given with the masculine singular agentive form of the adjective 'big'.

(5)	*Wali & Koul (1997)*	*Grierson (1911)*	*Bailey (1937)*
	bɜɖʲ	baɖⁱ	bɜɖi

Contradictory statements can be found in the literature on the question whether these 'palatalized retroflexes' in Kashmiri should be treated as phonetically and phonologically one segment or as a sequence of a retroflex with a short vowel or glide. Bhat (1987) remarks that the presence of "very short vowels" in contemporary Kashmiri has become obsolete, indicating that there is no reason (neither phonetically nor phonologically) to assume a sequence of two gestures or two phonemes. In his description of Kashmiri, Morgenstierne (1941) however points out that there are differences in the phonetic realization within the group of mātrā vowels. Whereas –u and – ü -mātrā (both causing velarization of the preceding consonant) are said to be inaudible by now in Morgenstierne (1941: 87), –i-mātrā still sounds like a very short [i], indicating a separate, additional i-gesture. As in the case of Toda, no articulatory data could be found for Kashmiri illustrating the actual realization of palatalized retroflexes, which could answer the question whether a simultaneous combination of retroflexion and palatalization gesture takes place.

The present study assumes that a combination of retroflexion and palatalization at the same time is impossible, and that the alleged patalalized retroflex segments in Toda and

[6] Spajić et al. (1996) do not mention what these speakers produce instead; it can be only speculated that these were non-palatalized retroflex rhotics.

Kashmiri have to be accounted for in another way. I propose that these segments are actually not retroflexes with a superimposed palatalized gesture, but phonetically sequences of a retroflexed segment followed by a short glide /j/. This proposal does not imply that the palatalized segment, which consists of two successive gestures, has to be phonologically interpreted as two phonemes instead of one. Toda and Kashmiri seem to be languages that chose to interpret the two gestures as belonging to one phoneme.

An indication that these gestures are indeed separate can be seen in the diachronic development of the assumedly palatalized retroflexes in Kashmiri. Diachronically, the mātrā vowels in Kashmiri stem from normal length vowels which have been shortened wordfinally (Morgenstierne 1941: 89). Kashmiri hence had two separate gestures which were assigned to different phonemes, a consonantal and a vocalic one. These gestures were categorized at a later stage as belonging to one phoneme.

Further evidence for the claim that there are two gestures instead of one can be found in the acoustic signal of the Toda trills (Spajić et al. 1996: 19). The signal for the palatalized trill /ɽʲ/ is 190 ms long, compared to the non-palatalized /ɽ/ which is 100 ms. The palatalized counterpart is thus nearly twice as long. There is no articulatory explanation why a palatalized segment should take longer to articulate than a non-palatalized one if one assumes that the two gestures co-occur together. If assuming however that two gestures are produced successively, the nearly double length of the palatalized segment compared to the non-palatalized is explainable.

In order to judge the values for palatalized and non-palatalized segment lengths, let us compare them to length measurements (in ms) of the palatalized and non-palatalized segment pairs in Russian from Bolla (1981), cf. the tables in (6) and (7), giving the labial articulations and the tongue dependent articulations, respectively. The labial articulations are discussed first, as they involve an independent secondary tongue gesture of palatalization.

(6)

	Labial					
	p	b	f	v	m	ratio
plain	116	120	128	115	97	1
palatalized	170	140	130	125	97	1.14

For this class, the length ratio of plain vs. palatalized is 1 : 1.14. This means that the palatal segment is on average a seventh longer than the plain segment. Thus, the secondary gesture seems to need some additional time to be articulated. It is striking that the values of all labial pairs lie rather close together apart from the voiceless stops, where [p] is 116 ms and [pʲ] is 170 ms long. Why this pair is departing so much from all the other labials cannot be answered here.

Velar and coronal articulations are expected to show roughly the same length for plain and secondary palatalized segments, because the palatalization is not an added gesture but changes the primary articulation (recall the description in 4.1). These expectations are fulfilled by the velar class (cf. left side of table (7)), where the palatalized signals are just slightly longer than the plain ones, as the ratio of 1: 1.08 indicates. The coronal class, however, departs from this picture; it has a ratio of 1 : 1.17, which is higher than that for the labial class. As both velar and coronal sounds undergo a change in primary articulation when palatalized, these results are unexpected. But looking at the coronal segments in detail we detect that this class does not behave homogenously. Whereas for the plosives and the voiced fricative the palatalized segment is longer than the plain one (the ratio varies there between 1 : 1.17 for the voiced plosive and 1: 1.66 for the voiced fricative), for the nasal and the voiceless fricative the plain signal is longer (average of 1 : 0.91). Again, no explanation for this behavior can be given.

(7)

	Velar				Coronal					
	k	g	x	ratio	t	d	n	s	z	ratio
Plain	169	150	110	1	134	102	105	195	96	1
palatalized	184	160	120	1.08	190	120	98	177	160	1.17

The Tamil retroflex rhotics have a ratio of 1 : 1.9 for plain vs. palatalized signal length, which is by far higher than any of the Russian ratios. This difference supports the claim made before that there might be two successive gestures involved in the articulation of the palatalized retroflex in Tamil, whereas palatalization of non-retroflex segments in Russian either is a simultaneous articulation of two gestures (as in the case for labials), or involves only one primary gesture, which differs from the non-palatalized counterpart (as in the case for coronals and velars).

Unfortunately, we do not have any further measurements for palatalized rhotics, so the present data are merely an indication that the hypothesis of successive instead of simultaneous gestures could hold. Further research has to be conducted on the exact articulation and gestural timing of palatalized rhotics in general and palatalized retroflex in particular. This may shed light on the articulatory timing of the gestures and further properties of their articulation.

5 Phonological representation

There is consensus in the phonological literature that retroflex segments are to be presented with the feature specification [coronal, -anterior, -distributed] (e.g. Chomsky & Halle 1968; Hume 1994; Hall 1997). The frequent co-occurrence of retroflexes with back vowels and their tendency to avoid front vowel context, as described e.g. by Bhat (1974) or Flemming (2001), lead several scholars such as Lin (1989) and Gnanadesikan (1994) to propose the addition of the dorsal feature-value [+back] in the representation of retroflexes. [+back] is phonetically defined as retraction. A complete representation of retroflexes with these features is given in (8).

(8)
```
                Place
              /       \
        Coronal       Dorsal
         /    \          |
   [–ant]  [–dis]    [+back]
```

Hall (1997, 2000), who assumes that Lardil has plain, i.e. non-velarized and non-palatalized, retroflex segments, proposes a presentation of retroflexes that can account for such segments and that differs slightly from that in (8). He introduces a three-way contrast of velarized, palatalized, and neither velarized nor palatalized retroflexes such as [ʈˠ], [ʈʲ], and [ʈ], respectively. All three classes share the place features [coronal, –anterior, –distributed]. They differ in the values that they are assigned for the dorsal feature [back]. Velarization is expressed by [+back], palatalization by [–back], and a plain retroflexion by [0back], for the last case the tongue back is assumed to be in neutral or unspecified position. This distinction is unnecessary if one assumes that all retroflexes are velarized, as it was argued for in the previous sections. Instead, the universal representation of all retroflex segments with the feature value [+back] is a phonological realisation of the phonetic implication made in (1), that all retroflexes are retracted.

The change of a retroflex segment into a laminal postalveolar in the process of secondary palatalization has to be phonologically represented as following. Laminal postalveolars are uncontroversally [coronal, –anterior, +distributed]. Furthermore, the domed, palatalized tongue dorsum of these segments has to be represented with a dorsal feature [–back]. The change of a retroflex to a laminal postalveolar in palatalizing front vowel context can then be accounted for by the feature [–back] spreading from the front vowel onto the retroflex (see e.g. Lin 1989 on the assumption that front vowels are [–back]). As a segment with the specification [coronal, –anterior, –distributed, –back] is phonetically not interpretable since [–distributed], i.e. apicals, have to be [+back], the palatalization process triggers a further change from [–distributed] to [+distributed]. The resulting segment, which is specified as [coronal, –anterior, +distributed, –back], has to be phonetically interpreted as a laminal postalveolar.

6 Conclusion

The present article took as starting point Bhat's (1974) study on the possible correlation of retroflexion and retraction. In his article, Bhat concluded "it is clear that retroflexion cannot be identified or correlated with retraction" (p. 237). It has been shown that the first of Bhat's observations, that retroflexion is not identical to retraction, is valid, as all apicals (and also velars) are retracted and thus retraction cannot be used as single defining property for retroflex segments. This problem can be solved by introducing further phonetic properties (as developed in Hamann in prep.) to define retroflexion. But opposing Bhat, it was illustrated that retroflexion always correlates with retraction, and that retroflex vowels in Badaga and retroflex consonants in Lardil are no counterexamples to this claim, since they are neither phonetically nor phonologically non-retracted.

From the correlation of retroflexion with retraction follows that retroflexes cannot be palatalized, because the gestures of retroflexion and palatalization are articulatory incompatible. It has been shown that there are alternative explanations for supposedly palatalized retroflexes in languages such as Toda and Kashmiri, but further research has to be conducted on these segments in order to support or disprove the hypothesis of successive retroflexion and palatalization gestures.

References

Bailey, T.G. (1937) The pronunciation of Kashmiri. London: RAS.
Bhat, D.N.S. (1974) “Retroflexion and Retraction.” Journal of Phonetics 2: 233-237.
Bhat, R. (1987) A descriptive study of Kashmiri. Delhi: Amar Prakashan.
Bolla, K. (1981) A Conspectus of Russian Speech Sounds. Budapest: Hungarian Academy of Science.
Brosnahan, L.F. & Bertil Malmberg (1970) Introduction to Phonetics. Cambridge: Heffer & Sons.
Borgstrøm, C.H. (1940) A linguistic survey of the Gaelic dialects of Scotland, vol. I: The dialects of the outer Hebrids. Norsk Tidsskrift for Sprogvidenskap. Suppl. Bind I. Oslo: Aschehoug.
Butcher, Andrew (1992) “The phonetics of neutralisation: the case of Australian coronals.” In: Jack Windson (ed.) Studies in general and English phonetics. London: Routledge; 10-38.
Catford, John C. (1977) Fundamental Problems in Phonetics.Edinburgh: University Press.
Ćavar, M. & S. Hamann (2002) “Polish velar and coronal palatalization – its perceptual background.” Ms. to be published in Proceedings of the Fourth European Conference on Formal Description of Slavic Languages.
Chomsky, Noam and Morris Halle (1968) The Sound Pattern of English. New York: Harper and Row.
Emeneau, M.B. (1939) “The vowels of the Badaga language.” Language 15 (1); 43-47.
Evans, Nick (1995) “Current Issues in the Phonology of Australian Languages.” In: John Goldsmith (ed.) The Handbook of Phonological Theory. London: Blackwell; 723-740.
Flemming, Edward (2001) Auditory Representations in Phonology. Routledge: London.
Gnanadesikan, Amalia (1994) “The Geometry of Coronal Articulations.” NELS 24,1: 125-139.
Grierson, George (1911) Standard manual of the Kashmiri language. 2 vols. Jammu: Light and Life.
Hall, T. Alan (1997) The Phonology of Coronals. Amsterdam: John Benjamins.
Hall, T.A. (2000) “Typological generalizations concerning secondary palatalization.” Lingua 110: 1-25.
Halle, Morris & Kenneth Stevens (1969) ‘On the feature “advanced tongue root.” Quarterly Progress Report 94. Cambridge, MA: Research Laboratory of Electronics, MIT, 209-215.
Hamann, Silke (in prep.) Phonetics and Phonology of Retroflexes. Ph.d. thesis manuscript of the University of Utrecht.
Hamilton, William S. 1980. Introduction to Russian Phonology and Word Structure. Columbus: Slavica.
Hume, Elizabeth (1994) Front Vowels, Coronal Consonants and their Interaction in Nonlinear Phonology. London: Garland.
Kelkar, A. R. & P. N. Trisal (1964) ‘Kashmiri word phonology: a first sketch.’ Anthropological Linguistics 6,1: 13-22.
Ladefoged, Peter (1971) Preliminaries to Linguistic Phonetics. Chicago: UCP.
Ladefoged, Peter & Peri Bhaskararao (1983) “Non-quantal Aspects of Consonant Production: A Study of Retroflex Consonants.” Journal of Phonetics 11: 291-302
Ladefoged, Peter & Ian Maddieson (1996) The Sounds of the World’s Languages. Oxford: Blackwell.
Laver, John (1994) Principles of Phonetics. Cambridge: CUP.
Lin, Yen-Hwei (1989) “The Retroflex as a Complex Segment.” ESCOL 6: 182-193.
Maddieson, Ian (1984) Patterns of Sounds. Cambridge: CUP.
Masica, Colin (1991) The Indo-Aryan languages. Cambridge: Cambridge University Press.
McCawley, James (1966) “Further revisions in Finnish rules.” Unpublished paper, University of Chicago.
Morgenstierne, Georg (1941) “The Phonology of Kashmiri.” Acta Orientalia 19,1: 79-99.
Scatton, Ernest A. (1975) Bulgarian Phonology. Columbus: Slavica Publishers.
Spajić, Siniša, Peter Ladefoged & P. Bhaskararao (1996) “The trills of Toda.” Journal of the International Phonetic Association 26,1: 1-21.
Spencer, Andrew (1984) “Eliminating the feature [lateral].” Journal of Linguistics 20: 23-43.
Stevens, Kenneth, Samuel Keyser and Haruko Kawasaki (1986) “Toward a phonetic and phonological theory of redundant features.” In: J.S. Perkell and D.H. Klatt (eds.) Invariance and Variability in Speech Processes. Hillsdale: Erlbaum. 426-449.
Stevens, Kenneth (1998) Acoustic Phonetics. Cambridge, Massachusetts: MIT Press.
Trask, R.L. (1996) A Dictionary of Phonetics and Phonology. London: Routledge.
Wali, Kashi & Omkar N. Koul (1997) Kashmiri: A cognitive-descriptive grammar. London: Routledge.
Wierzchowska, Bożena (1971) Wymowa Polska. Warzawa: Państwowe Zakłady Wydawnictw Szkolnych.
Wilkinson, Karina (1988) ‘Prosodic Structure and Lardil Phonology.’ Linguistic Inquiry 19,2: 325-334.
Wood, Sidney A.J. (1996) “Assimilation or coarticulation? Evidence from the temporal co-ordination of tongue gestures for the palatalization of Bulgarian alveolar stops.” Journal of Phonetics 24: 139-164.

The Interaction of Syllabification and Voicing Perception in American English

Kenneth de Jong, Kyoko Nagao, and Byung-jin Lim
Department of Linguistics, Indiana University

1 Introduction

1.1 Two ways of explaining phonology by phonetic facts

Being scholars who have been experimentally investigating language phenomena within the discipline of linguistics, it is heartening to see phonological theory over the past 50 years incorporating more and more physical attributes of the speech communication process into the explanation of phonological phenomena. As, we hope, interest continues in seeing what aspects of phonetic behavior can account for what aspects of phonological systems, it will soon become apparent that there are a wide variety of ways this can be done. In particular, there are two very different types of approaches which are often not clearly distinguished in discussions of phonetic explanation.

The first type explains parts of previous phonological analyses as being due to some specified phonetic strategy. Often, attributes previously accounted for in a phonological grammar may be accounted for by a specified set of production strategies that speakers use to accomplish a more sparsely specified phonological structure. A good example of this is in Pierrehumbert & Beckman's (1988, also Beckman & Pierrehumbert, 1986) model of Japanese tonal structure. Most phonological descriptions of Japanese tone structure previous to their work posited tonal specifications on each mora, along with context sensitive rules to account for, among other things, the apparent lack of an initial low tone in words with an initial accent. In Pierrehumbert & Beckman's model, the phonological specification of tone is more sparse and regular, the apparent loss or adjustment of tones in various contexts is due to the phonetic spell out of these tones. A conventionalized phonetic execution of a slimmed down phonological structure accounts for much of what one observes. This sort of approach does not require an emphasis on production. Kingston & Diehl (1994) employ a similar general strategy, though in their model the production strategies are heavily driven by the expected needs of the perceiver. What these models have in common with one another is 1) the on-line phonetic skills of the speaker explain phonological phenomena, and 2) the phonetics is either smart (as in Pierrehumbert & Beckman), or very smart (as in Kingston & Diehl).

The second type of approach, while perhaps not employed as commonly in current research, also has a very long track record. In this approach, a phonological phenomenon is due to very low-amplitude phonetic pressures constantly acting on the communication process over an extended period of time. Here, phonetic facts take a long time to create a phonological phenomenon. What makes this approach different from that outlined above is that the phonetic skills in individuals may be essentially irrelevant. It is historical dynamics which does the work of putting phonetic facts into the phonological grammar. Two excellent examples of this sort of model are Quantal Theory (Stevens, 1989) and 'Adaptive Dispersion Theory' (Liljencrants & Lindblom, 1971; Lindblom, 1986, Lindblom & Maddieson, 1988;

Lindblom & Diehl, 2001). Both of these models attempt to explain typological phonological facts in terms of phonetic facts, non-linearity in the 'links of the speech chain', and the shape of the articulatory space modulated by the auditory system, respectively. Note that, while attempts have been made to determine how such phonetic facts impact the phonetic skills of individuals (such as Perkell & Nelson, 1985, Perkell & Cohen, 1989; and Beckman et al., 1995, for Quantal Theory), such models do not need the individual phonetic skills to be demonstrably constrained by the phonetic factor which is doing the work. Also note that, while both models are formulated to explain typological facts, they are both amenable to explaining particular phonological facts of particular languages, as is demonstrated by de Jong & Obeng (2000).

Distinguishing between these two types of models is particularly important, then, since the predictions made by any particular model with respect to how any given individual will behave in an experimental setting depends on which view of the phonetic incorporation it embodies. Also, we cannot properly say we have understood a phonological phenomenon unless we can actually link the part of the phonology we are attempting to explain to the phonetic factor. Hence, with the second type of model, it is also necessary to understand how the phonetic factors actually give rise to the historical dynamics which produce the phonological phenomenon.

1.2 Syllable level voicing allophony

The current paper explores these two sorts of phonetic explanations of the relationship between syllabic position and the voicing contrast in American English. It has long been observed that the contrast between, for example, /p/ and /b/ is expressed differently, depending on the position of the stop with respect to the vowel. Preceding a vowel within a syllable, the contrast is largely one of aspiration. /p/ is aspirated, while /b/ is voiceless, or in some dialects voiced or even an implosive. Following a vowel within a syllable, both /p/ and /b/ both tend to lack voicing in the closure and the contrast is expressed largely by dynamic differences in the transition between the previous vowel and the stop. Here, vowel and closure duration are negatively correlated such that the /p/ has a shorter vowel and longer closure duration. This difference is often enhanced by the addition of glottalization to /p/. In addition to these differences, there are additional differences connected to higher-level organization involving stress and feet edges. To make the current discussion more tractable, we will restrict ourselves to the two conditions (CV and VC) laid out above.

The most straightforward traditional phonological approach to these facts would be simply to specify a context sensitive rule which either changes the identity of the two stops to something more closely approximating what appears phonetically, or to have language specific realization rules affecting rendering of some underlying feature, say [voice].

However, there are two facts which point out the shortcomings of this approach. 1) the type of allophonic variation we get in American English is also attested in a number of different languages, as pointed out by Keating, Linker & Huffman (1983). The pattern seems to be more general than would be predicted by an array of language-specific rules. 2) the allophonic difference between stops of the same type in different location seems to reflect a more general property of consonants in those syllabic positions. As is shown in, e.g., Turk (1993); de Jong (1998, in press), and others, consonants in prevocalic position tend to be more extreme in articulation and more temporally compact than consonants in post-vocalic position. Hence, expressing the voicing contrast in VC position in terms of vowel dynamics rather than in terms of contrasts linked to the consonant closure seems to 'make sense'. A general property of consonants in that location is encoded in the allophonic variants of the stops.

Jakobson, Fant, & Halle (1952) have a brief discussion of the relationship between the voicing allophony and syllabic position which captures this non-independence. In their discussion of Danish, they state that /t/ is 'strong' and /d/ is 'weak'. Here, other features, such as aspiration and voicing are redundantly specified to enhance the main contrast. Syllable positions also are 'strong' or 'weak'. "Two positions are discernible in the Danish word - strong and weak. In monosyllabic words the strong position for a consonant is at the beginning of the syllable and the weak position, at its end." (p. 5) Hence onsets and /t/ have an affinity, while codas and /d/ also have an affinity. Allophonic interaction between the two then is simply expressed as an additive relationship between the two factors, as illustrated in Figure 1.

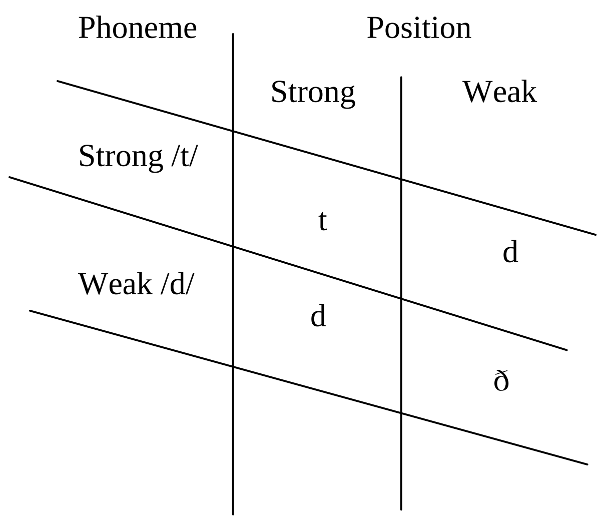


Figure 1. A schematic of the interaction between syllable position and the stop voicing contrast. (After Jakobson et al., 1952, Figure 1.)

Critical for this explanation is the existence of syllabic locations as having physical properties, here 'strong' and 'weak'. Thus, this sort of explanation runs counter to a trend in phonological theory toward explaining syllable-level effects as being due to some aspect of segmental sequencing, and not due to syllable level prosody at all. One such model is that of Kawasaki (1982) and Ohala (1990), who propose that syllable-level sequencing constraints are the result a long-term pressure toward acoustic modulation between successive segments. Similarly Mattingly (1984) notes that collation of segments at the level of the syllable tends to maximize the use of the frequency range used for speech transmission. Hence, any pressure toward long-term optimization of the transmission will tend to produce syllable level sequencing constraints. More recently, others, such as Wright (1996), Silverman (1996), and Steriade (1999), have articulated models in which specific syllable-level effects are due to the demands preserving specific segmental contrasts. In each of these models, the syllable is essentially an epiphenomena, and the syllable level phenomena are to be explained in terms of segmental contrast and sequencing.

Along these lines, Silverman (1998) presents a brief analysis of the voicing allophony in American English. The nature of this analysis as indicating a purely historical analysis in which syllable-level organization plays no part is spelled out in a more recent development of this work (Silverman, ms). In this analysis, he adopts Jakobson et al.'s (1952) description of the contrast between /b/ and /p/ as one of strength; /b/ is lenis, and /p/ is fortis. The syllable-conditioned allophony is generated by mechanisms specific to pre-vocalic and post-vocalic sequencing. In the CV situation, there is a historic tendency to shifting the lenis consonants toward voicelessness, and a subsequent subtle perceptually-driven bias applied to the distribution of fortis consonants which tends to separate them from the now voiceless lenis stops. The VC situation is somewhat different in that the response to a historic tendency to push both the fortis and lenis consonants toward voicelessness is a switch to a perceptually more salient dimension involving the dynamics of the stop closing movement. Fortis consonants involve relatively fast closing movements and lenis consonants involve relatively slow closing movements. This contrast is subsequently enhanced by the addition of glottal stoppage in the fortis case, further shortening the acoustic transition.

In this analysis, Silverman repeatedly makes the point that allophonic variation is not generated by speakers optimizing their production system, but rather is due to low-level, perhaps undetectable biases in the transmission of speech patterns from one generation to another. Here, speakers simply copy the distributions of phones from previous generations. Syllable structuring is irrelevant for the allophony. Perceptual selectivity tends to produce certain types of effects which are sensitive to the temporal sequencing of sounds, and so can *post-hoc* be abstracted into syllabic units. However, one does not need syllable-level parsing to explain syllable-level allophony.

2 The current study

2.1 Rationale for the experimental approach

The current paper examines the degree to which synchronic aspects of the speech communication system can be implicated in the appearance of syllable-level voicing allophony in American English. Simply put, if the phonological facts of syllable-level allophony are to be explained by phonetic facts, are such facts detectable in the production and perception system of individual speakers and hearers? Or do phonetic facts create allophony only through long-term historical dynamics which are not robustly detectable in experiments on individual speakers?

The current project probes the relationship between voicing[1] and syllable position by systematically controlling rate in a repetitive speech production task. This variable and this task were chosen due to an observation by Stetson (1951, and documented by Tuller and Kelso, 1991) that syllabic parsing is sensitive to rate of repetition. Stetson noted that fast rates induce perceptual resyllabification of codas (such as the /b/ in 'eeb') into onsets (as in 'bee'). Hence, this task allows us to investigate how consonants get parsed into syllables. By including consonant voicing differences, we can investigate the relationship between syllabic parsing and the voicing category of the consonant being parsed.

Thus, our inquiry into the occurrence of syllable-level allophony is indirect in two ways, in terms of technique, and in terms of specific questions posed of the data. First, in terms of technique, the project includes both a production component and a perception component. Speakers produce various forms with variation controlled explicitly in a production experiment. Their productions, then, are evaluated not in terms of their physical

[1] Throughout the description of the current data set, we will use the term 'voicing' to indicate the contrast between /p/ and /b/. This is purely a convenience and is not meant to indicate that the contrast is actually expressed or perceived on the basis of the occurrence of voicing in the stop closures. We will use the terms 'fortis' and 'lenis' when discussing interpretations of the data which implicate a model such as Jakobson, et al.'s or Silverman's.

make-up, but rather in terms of the category structure of perceivers. By evaluating productions in terms of perceptions, this study allows us to examine the degree to which the production and the perception system match one another, and also allows us to examine how the control variables in the production experiment affect the relationship between production and perception.

Second, in terms of questions posed of the data, we will pursue the question of how allophonic differences in stops and syllabic parsing affect one another. One type of hypothesis would be that they do not affect one another at all, speaker productions simply match perceptions. This would be the case if there is essentially no on-line relationship between allophony and syllabic location; speakers just produce items which correspond to the categories listeners have. However, another possibility is that there is a systematic misalignment between the productions and perceptions due to some production factor associated either with syllabic position or voicing which tends to create mismatch between production and perception. These mismatches would tend to push the phonological categories in particular directions over the long run.

2.2 Methods

Stimuli for the perception experiment were obtained by means of a production experiment involving four native speakers of some variety of mid-western American English. There were two female and two male speakers, all in their 30's, one of which is the primary author. Speakers were given one of four visual stimuli, 'eep', 'pea', 'eeb', or 'bee', and then asked to produce the form one time for each click of a metronome. A metronome was used to control the rate of repetition in order to insure a wide range of rates which are comparable across the four speakers. The stimuli used in the perception experiment were extracted from a condition in which the metronome started at a slow repetition rate of 450 ms/syllable, and which increased in rate by 12.5 ms increments until reaching a fast rate of 200 ms/syllable. This task did not prove difficult for the speakers, who performed it with little practice and with only a very small number of dysfluencies. In addition, previous analyses of the data show that speakers were quite good at tracking the rate of the metronome with their speech. Detailed analyses of the acoustics of these productions related to syllabification and rate changes are reported in de Jong (2001a) and de Jong (2001b), respectively.

The current paper, however, examines perceptual responses to these productions. In order to get judgements of voicing and syllabification at times throughout the utterances, three-syllable pieces were spliced from the original utterances, so that each syllable in the utterance was the middle syllable of one of the three-syllable stimuli. Two different splicing techniques were used, one which included consonantal transients and one in which vocalic portions formed the edge of the stimuli. Responses for the two techniques were virtually identical, indicating that listeners were very good at abstracting away from the spliced edges of the stimuli. (These results are reported in more detail in de Jong et al., 2001). Results presented below are for tokens with consonant transients included. Only the last 23 syllables were included, thereby eliminating the first syllables which tended to vary while the subjects entrained to the metronome. Thus, there were 21 stimuli for each utterance, one utterance for each speaker by word form were included for a total of 4 speakers X 4 word forms X 21 stimuli = 336 stimuli.

These stimuli were then presented to 23 native listeners of American English by means of a *Matlab* protocol on PC workstations. Listeners were presented with a stimulus, told that the sounds were spliced out of a longer utterance in which a speaker was repeating a syllable, and asked to identify the repeated syllable. They were given the same four choices as the speakers were presented with originally, 'eep', 'pea', 'eeb', and 'bee'.[2] The task was self-paced, and the listeners could repeat each stimulus as many times as they wished before making a judgement.

2.3 Predictions concerning perceptual identification

With respect to the syllabification of the stimuli, we expected, based on informal evaluation, that listeners would perceive fast rate VC's as being CV's, just as Tuller & Kelso had found. However, the point of the current analysis is if and to what degree the voicing of the consonant affects this pattern of syllabic judgement. Analyses such as Jakobson et al.'s seem to predict that this perception of VC's as CV's should be encouraged by the consonant being a fortis /p/. Hence, 'eep' should be more likely to be identified as a CV than 'eeb'. If this effect is strong enough, it might also even induce a misperception in the opposite direction, such that a lenis /b/ in onset position might tend to be misidentified as a VC form. In contrast to this, an explanation of allophony such as Silverman's would predict no necessary effect of voicing on syllabification, since allophony is only indirectly related to syllabic parsing.

With respect to voicing judgements, the same sorts of predictions can be made concerning the syllabic location of the consonants. Again, analyses such as Jakobson et al.'s seem to predict that CV's would tend to be identified as fortis /p/, while VC's would tend to be identified as lenis /b/. A purely historical account, especially one in which listener's perceptions are based on an internal model of the distribution of productions would seem to predict no relationship between syllabification and voicing. However, a complication here is that historical accounts such as Silverman's rely on various phonetic effects of markedness to drive the changes which lead to allophonic variation. Specifically, lenis /b/ occupies the unmarked voiceless position, while fortis /p/ was driven away from the unmarked position by means of perceptual selectivity. This asymmetry would suggest that fortis /p/'s would tend to appear in the unmarked position, especially when the effects of the production factor (here, speech rate) are strong. Hence, one might find CV's would tend to be identified as lenis /b/, the category which occupies the unmarked position. It is unclear whether such an argument can be made for VC's since the markedness of the dynamics of the voicing effect in the vowel is unknown.

3 Results

3.1 Voicing effects on syllabic parsing

Rate effects on the perception of syllabic affiliation are presented in Figure 2. Here, the horizontal axis plots the position of the stimulus in the original utterances, and hence is an index of speech rate. Stimuli with low numbers appeared closer to the beginning of the utterance, and hence had slower rates than did ones later in the utterance appearing to the left in the figure. The vertical axis plots the proportion of CV identifications by all listeners combined.

[2] The listeners were also asked to give an indication as to how confident they were in their assessment by clicking on a scroll-bar, but these results are not considered in the current paper. More on this is presented in Nagao, et al. (2001).

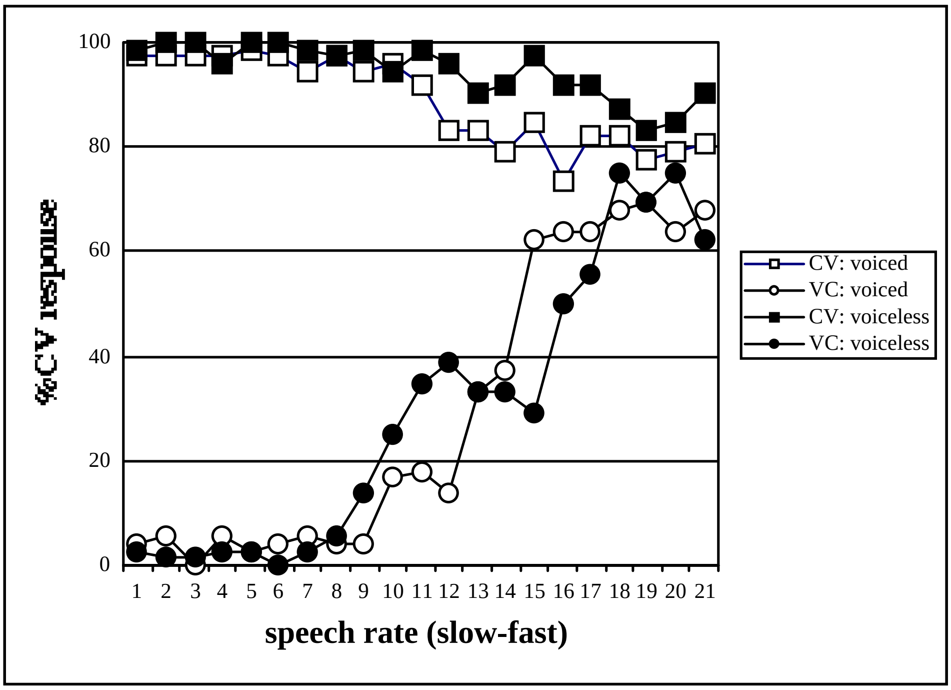


Figure 2. Proportion of CV responses as a function of position in utterance which acts as an index of rate. Functions given here indicate responses for the four different produced syllables.

The general pattern of results for VC productions (circles in Figure 2) is what is expected from Stetson's and Tuller & Kelso's observations. VC productions are consistently identified as VC forms when repeated at slow rates, as is evident in the functions lying very close to zero. As rate increases, however, one sees a sharp increase in CV identification, trailing off at a level of approximately 75% at the fastest rates. There is an additional aspect of the current results which is not mentioned in previous works, which is a complementary shift which happens with CV productions (squares in Figure 2). At slow rates, CV productions are identified as such, the functions lying very close to 100%. As rate increases, there is a smaller downward shift which occurs at roughly the same location in the utterance, again to a level of approximately 75%. Analyses of this effect are currently under way; preliminary results suggesting that part of this effect is due to a generally greater degree of difficulty in identifying fast rate tokens. However other aspects of the analyses suggest that this effect is due to genuine ambiguity of the tokens with respect to whether the consonants belong with the previous or following vowel.

What is particularly of interest in the current discussion, however, is the difference between the voiced and voiceless forms (filled and hollow symbols, respectively). There is essentially no difference between the two sets of functions. Both sets show the same shifting pattern at the same locations in the utterances and to the same degree. There are subtle differences in the two functions, especially in the middle portions of the utterances; however, these differences do not seem to reflect a consistent bias in one direction or another. Some voiceless forms show more CV identifications, while others show less. In addition, an increment in proportion of CV responses for VC productions tends to get counteracted by a greater decrement for CV productions. The overall result of all this is that there is no discernable pattern of voicing effect on perceived syllabification.

To quantify these observations, we conducted a series of regression analyses, linking production types and rate variation to acoustic attributes and then to identification responses. Table 1 summarizes the relationship between the production control variables, voicing, syllabification, and rate, and perceived identification. As expected, most of the perceived syllabification in the corpus corresponds to the stimulus given to the speaker, while a smaller but still substantial proportion of perceived syllabification is predicted by rate increases. The proportion accounted for by consonant voicing is remarkably small, less than one tenth of a percent of the variability in perceived syllabification.

Table 1. Summary of regression of proportion CV responses against production control variables.

	R^2	*p*
Syllabification (Intended)	.633	<.001
Rate	.061	<.0001
Voicing	.001	.503

As predicted by a model such as Silverman's, syllabic parsing is not affected by the voicing of the consonant being parsed. Thus, while the location of the consonant with respect to a vowel may be historically responsible for a shift in the phonetic nature of the consonant, these differences in the phonetic nature of the consonant are not used to parse the location of the consonant with respect to the vowel.

3.2 Syllabification effects on voicing identification

Figure 3 is similar to Figure 2, except that the vertical dimension here is proportion of voiceless responses rather than proportion of CV responses. Also, only CV productions are plotted. The general pattern with respect to rate is quite similar to that presented in Figure 2; voicing identification neatly matches the stimulus presented to the speakers at slow rates, but stops matching at faster rates. Specifically, there is a fairly large effect of produced voiced syllables ('bee' productions) being identified as voiceless ('pea'). Hence, there is a clear effect of rate inducing a shift in the productions which crosses the category boundary between /p/ and /b/.

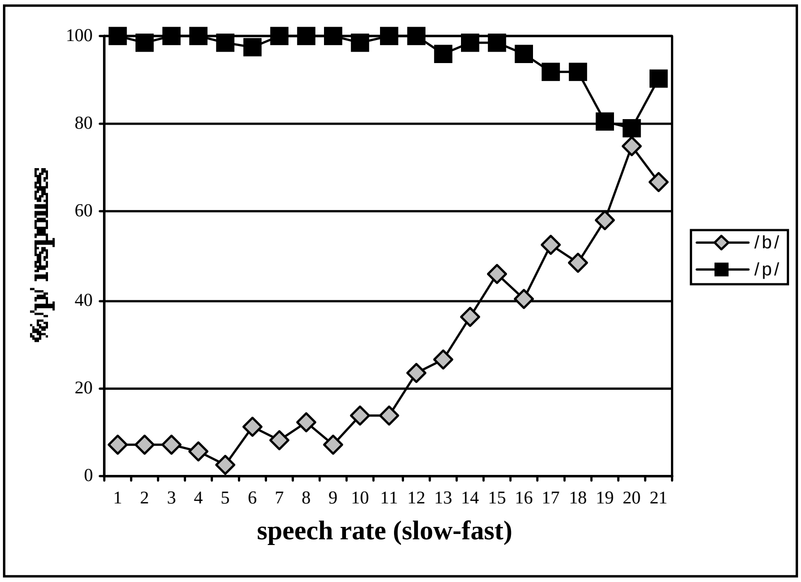


Figure 3. Proportion of voiceless (/p/) responses as a function of position in utterance which acts as an index of rate. Functions given here indicate responses for onsets only.

Figure 4 is the same as Figure 3, except that it plots VC productions rather than CV productions. Here we also see a mismatch between production and perception, however, in the opposite direction. Here, voiced productions are consistently identified as voiced, while voiceless productions tend to be misperceived as voiced. A rate effect is evident here as well, however it is more subtle, and tends to be continuous across all of the rates.

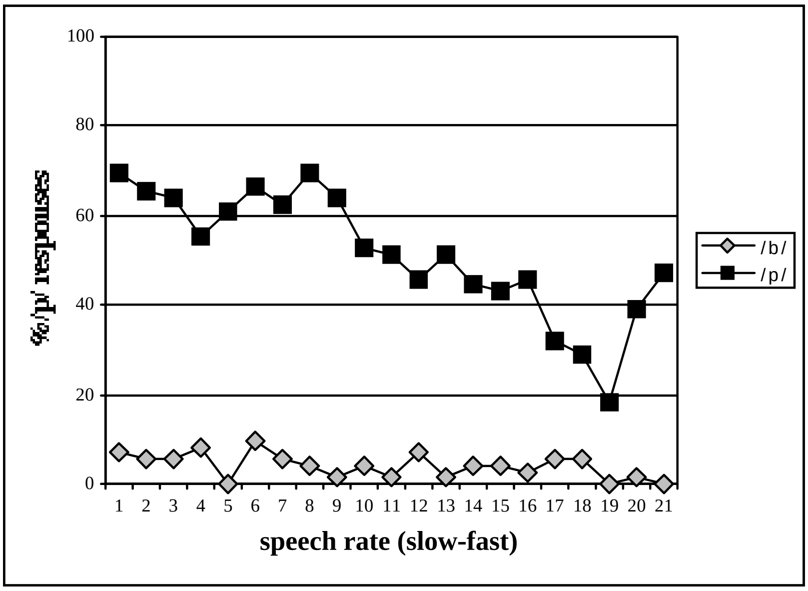


Figure 4. Proportion of voiceless (/p/) responses as a function of position in utterance which acts as an index of rate. Functions given here indicate responses for codas only.

Thus, it appears that some of the expectations based on Jakobson et al.'s analysis are born out in the current data. Onsets tend to be identified as voiceless, while codas tend to be identified as voiced. To illustrate this point, Figure 5 compares the schematic from Figure 1 to average data from the current experiment. Here, averages are calculated for the four stimuli presented to the speakers. Averages are in terms of proportion of times in which an item is identified as voiceless. As can be seen, the relationship apparent in the data to the right in Figure 5 is essentially that schematized to the left in Figure 5. What is interesting is that the vertical axis of the schema is labeled as 'strength', while the vertical axis in the data figure is proportion of /p/ responses. Thus, here, 'strength' is indexed in the likelihood that a perceiver will label a consonant as being a /p/. Hence, /p/'s in onsets are better /p/'s than ones in codas, and /b/'s in codas are better /b/'s than ones in onsets. Or, to restate Jakobson et al.'s analysis, being in an onset and being produced as a /p/ correlate, and being in a coda and being produced as a /b/ also correlate in the opposite direction.

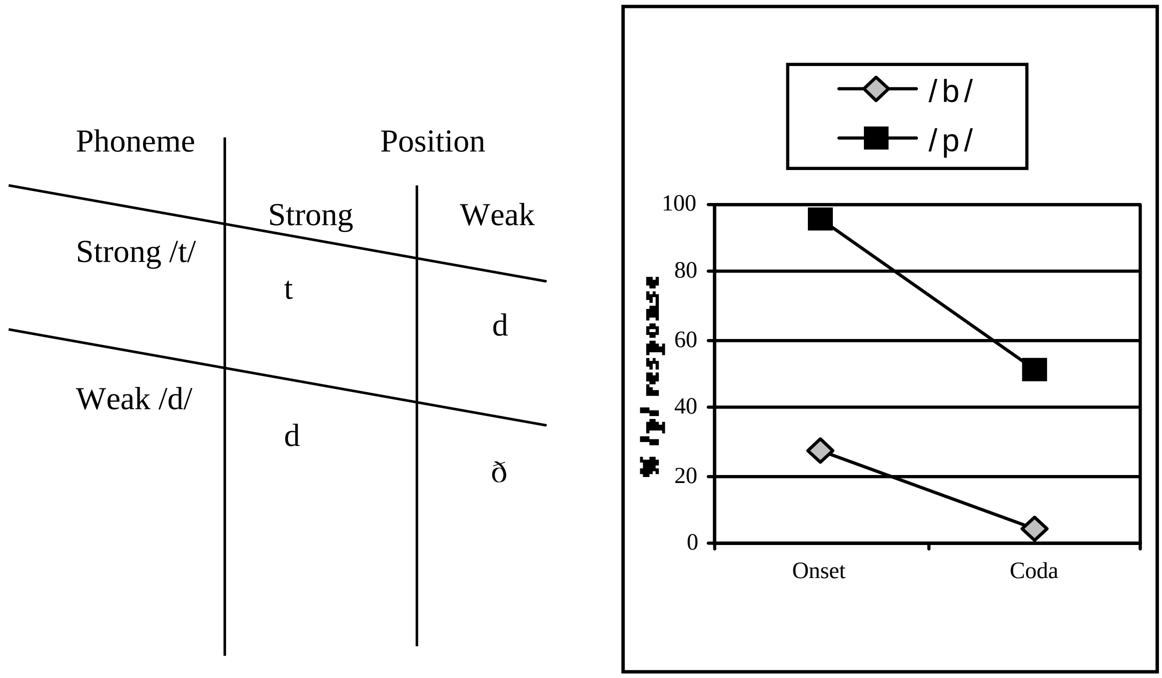


Figure 5. Average proportion of identifications as voiceless for each stimulus presented to the speakers. The schema from Figure 1 is repeated here for comparison.

4 Discussion and conclusions

At first glance, it appears that the current study has produced inconsistent results with respect to the general models being tested here. Stating these results with respect to our extension of Jakobson et al.'s analysis of Danish makes the point quite well. We predicted on the basis of this analysis that /p/'s make better onsets and /b/'s make better codas, and hence should tend to bias the identification of syllabic structure accordingly. This proved to be wrong; no bias is evident in the current data. We also predicted that onsets should make for better /p/'s and codas should make for better /b/'s. This prediction was born out. Hence, there is an interaction between location of the consonant and its identification, while there is no interaction between the identity of the consonant and its perceived location.

What these results suggest is a fundamental difference between the two identification tasks with respect to the production task. The syllabification task is a parsing judgement concerning the grouping of the material in the speech signal. It is heavily affected by rate scaling, a task which, according to detailed analyses of the current corpus in de Jong (2001a, and 2001b) has a disproportionate effect on the portion of the signal at the edge of each syllabic constituent. The identity of the consonant does not heavily affect the acoustics of these edges; both voiced and voiceless consonants have roughly the same global temporal structure. The voicing identification task, however, is a paradigmatic categorization task which relies on a wide range of much more subtle information. The difference in the robustness of the information on which these judgements rely has been implicated in a study which is parallel to the current one involving non-native listeners (Lim, de Jong, & Nagao, in press; Nagao, de Jong, & Lim, 2001). Non-native listeners were speakers of Korean and Japanese, two languages which are notably different from English in both the voicing categories and in syllabic inventory. While the non-native listeners show a number of differences from native listeners in voicing categorization, they are remarkably similar to natives in their performance on syllabification judgement. What this suggests is that the parsing task is easier than the categorization task.

Hence, to reevaluate the conclusion reached above that the lack of an effect of voicing on syllabification supports a purely historical model of allophony, it is apparently the case that syllable level allophony is not there to help with syllabic parsing. This is consistent with a historical model of the allophony. However, it is also not the case that syllable level allophony is not synchronically connected to syllabic organization. Being produced in a certain syllabic position does, apparently, affect the production of voicing cues in the signal. Being produced as an onset tends to make a stop fortis, while being produced as a coda tends to make a stop 'lenis'. The direction of these effects is what would be expected of the pressure which would have driven the allophonic variation observed in the current system. That is, it is in the direction suggested by Jakobson et al's analysis. In addition, increasing rates tends to make this effect larger, suggesting that there is some sort of production factor underlying the effect.

In conclusion, diachronic modelling of the segmental phonetic pressures has some merit for explaining syllable-level allophony. However, the current study suggests that the pressures are not just apparent in a subtle generation-to-generation shift, detectable only through sampling entire production distributions. They are actually detectable in an experimental study which examines the relationship between production and perception under systematically varied conditions. In the current study, the production task employed here induces the appropriate changes, when evaluated with respect to the listeners' perceptual systems. That is to say, examining the relationship between production and perception in a communicative system, under conditions of controlled variation, gives a an indication of the sort of disequalibria which would over the long term tend to push phonological systems in particular directions. What this points out is that even long-term historical models are subject to experimental investigation, and that, within such models, the phonetics of individuals contain the structure which are found in phonological conventions.

References

Beckman, M.E., & J.B.Pierrehumbert (1986). Intonational structure in English and Japanese. *Phonolotgy*,3: 255 - 309.

Beckman, M.E., T.-P. Jung, S.-H. Lee, K. de Jong, A.K. Krishnamurthy, S.C. Ahalt, K.B. Cohen, & M.J. Collins (1995). Variability in the Production of Quantal Vowels Revisited. *Journal of the Acoustical Society of America,* **97**: 471 - 490.

de Jong, K.J. (1998). Stress-related variation in the articulation of coda alveolar stops: Flapping revisited. *Journal of Phonetics***,** **26**: 283 - 310.

de Jong, K.J. (in press). Temporal constraints and characterizing syllable structuring. In. J. Local, R. Ogden, & R. Temple (eds.), *Papers in Laboratory Phonology VI*. Cambridge: Cambridge University Press.

de Jong, K.J., B.J. Lim, & K. Nagao (2001). The Perception of Rate Induced Resyllabification in English. (Abstract) *Journal of the Acoustical Society of America*, 109: 1pSC4.

de Jong, K.J., & S.G. Obeng (2000). Labio-Palatalization in Twi: Prosodic, contrastive, and quantal effects on linguistic micro-evolution. *Language*, 76: 174 - 195.

Jakobson, Fant & Halle (1952). *Preliminaries to Speech Analysis*. Cambridge, MA: MIT Press.

Kawasaki, H. (1982). *An Acoustical Basis for the Universal Constraints on Sound Sequences*. Ph.D. dissertation, University of California, Berkeley.

Keating, P.A., W. Linker, & M. Huffman (1983). Patterns in allophone distribution for voiced and voiceless stops. *Journal of Phonetics*, 11, 277 - 290.

Kingston, J., & R.L. Diehl (1994). Phonetic knowledge. *Language*, 70: 419 - 454.

Liljencrants, J., & B. Lindblom (1972). Numerical simulation of vowel quality systems: the role of perceptual contrast. *Language*, 48: 839 - 862.

Lim, B.J., K.J. de Jong, and K. Nagao (in press). Cross-language perception of syllable affiliation: effects of voicing and language background. *Berkeley Linguistic Society,* 26.

Lindblom, B. (1986). Phonetic universals in vowel systems. In J.J. Ohala and J.J. Jaeger (eds.), *Experimental Phonology*. Dordrecht: Foris.

Lindblom B., & I. Maddieson (1988). Phonetic universals in consonant systems. In L.M. Hyman, and C.N. Li (eds.), *Language, Speech and Mind*. London: Routledge.

Linblom, B., & R.L. Diehl (2001). Reconciling static and dynamic aspects of the speech process. (Abstract) *Journal of the Acoustical Society of America*, 198: 2380, 2pSC1.

Mattingly, I. (1981). Phonetic representation and speech synthesis by rule. In T. Myers, J. Laver, & J. Anderson (eds.), *The Cognitive Representations of Speech,* pp. 415 - 420. Amsterdam: North-Holland.

Nagao, K., B.J. Lim, and K.J. de Jong (2001). Cross-language Perception of Rate Induced Resyllabification. (Abstract) *Journal of the Acoustical Society of America*, 109: 4pSC12.

Ohala, J.J. (1990). Alternatives to the sonority heirarchy. In M. Ziolkowski, M. Noske, & K. Deaton (eds.), Papers from the 26th Regional Meeting of the Chicago Linguistics Society: Vol. 2: The Parasession on the Syllable in Phonetics and Phonology. Chicago: Chicago Linguistics Society.

Perkell, J.S., & M.H. Cohen (1989). An indirect test of the quantal nature of speech in the production of the vowels /i/, /a/ and /u/. *Journal of Phonetics*, 17: 123 - 133.

Perkell, J.S., & W.L. Nelson (1985). Variability in the production of the vowels /i/ and /a/. *Journal of the Acoustical Society of America*, **77**, 1889 - 1895.

Pierrehumbert, J.B., & M.E. Beckman (1988). *Japanese Tone Structure*. Cambridge, MA: MIT Press.

Silverman, D. (1996). Phonology at the interface of phonetics and morphology: Root-final laryngeals in Chong, Korean, and Sanskrit. *Journal of East Asian Linguistics,* 6: 1 - 22.

Silverman, D. (ms). English alveolar stops, and the nature of allophony, University of Illinois at Urbana-Champaign.

Steriade, D. (1999). Alternatives to the syllabic interpretation of consonantal phonotactics. In O.Fujimura B.Joseph and B.Palek (eds.) *Proceedings of the 1998 Linguistics and Phonetics Conference,* pp. 205-242. The Karolinum Press.

Stetson, R. H. (1951). *Motor Phonetics*. Amsterdam: North-Holland

Stevens, K.N. (1989). On the quantal nature of speech. *Journal of Phonetics*, 17: 3 - 45.

Tuller, B., & Kelso, J. A. S. 1991. The production and perception of syllable structure. *Journal of Speech and Hearing Search*, 34: 501-508.

Turk, A (1994). Effects of Position-in-Syllable and Stress on Consonant Articulation, Ph.D. dissertation, Cornell University.

Wright, R.A. (1996). Consonant Clusters and Cue Preservation in Tsou. Ph.D. dissertation, UCLA.

Subphonemic and Suballophonic Consonant Variation: The Role of the Phoneme Inventory

Lisa M. Lavoie
Research Laboratory of Electronics, Speech Communication Group
Massachusetts Institute of Technology
lisa@speech.mit.edu

Abstract

Consonants exhibit more variation in their phonetic realization than is typically acknowledged, but that variation is linguistically constrained. Acoustic analysis of both read and spontaneous speech reveals that consonants are not necessarily realized with the manner of articulation they would have in careful citation form. Although the variation is wider than one would imagine, it is limited by the phoneme inventory. The phoneme inventory of the language restricts the range of variation to protect the system of phonemic contrast. That is, consonants may stray phonetically into *unfilled* areas of the language's sound space. Listeners are seldom consciously aware of the consonant variation, and perceive the consonants phonemically as in their citation forms. A better understanding of surface phonetic consonant variation can help make predictions in theoretical domains and advances in applied domains.

1 Consonant variation

The consonant variation of interest here goes beyond well-recognized allophonic variation. This variation is not necessarily conditioned by position in word, position in syllable or by segmental context as allophonic variation is, but rather by other factors. In several areas, work has been done on variation that can be considered suballophonic, as opposed to strictly allophonic. Consonants have been found to respond phonetically to a range of conditioning environments, with: hyperarticulation under stress (De Jong 1995), reinforcement in prominent prosodic positions (Fougeron 1998, Fougeron & Keating 1997), reduction in colloquial speech (e.g., Brown 1990, Kohler 1990), and reduction in highly predictable words (Jurafsky et al. 1998). Even if these kinds of phonetic differences are noticed, they are seldom considered part of a language's phonology proper. Many factors, above and beyond those just mentioned, impinge upon a consonant's realization. This paper presents data that illustrate the role of the language's phoneme inventory in constraining some of that variation.

1.1 Instrumental work

Because so much consonant variation escapes the ear, instrumental work is essential to gathering accurate data. Lavoie (2001) located much more variation in consonants than was expected based on the available literature. Lexical knowledge and categorical perception both make it difficult to hear all of the variation that is present in consonants. Barry (1996, p. 115) explains the need to examine spectrograms, in addition to listening and transcribing the speech.

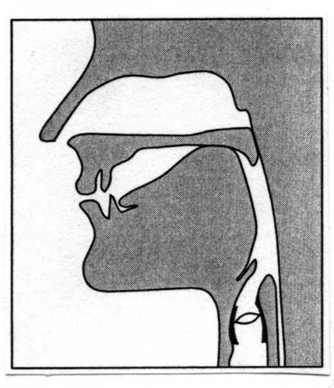
> *[…], the visual scrutiny of graphically presented instrumental analyses, particularly the use of good spectrograms as an accompaniment to careful auditory examination can serve to augment (and correct) the auditory analysis of even the most expert transcriber.*

One case where spectrograms are nearly essential is the case of unreleased stops. With a spectrogram, the presence or absence of a release burst is easily verified. Bursts are not always detectable when listening to continuous speech. Missing the burstless stops is likely due to the ways in which linguists are trained to identify phonetic segments and assign them symbols, which focuses on elements which are phonemic in some language. Since burstless stops are never phonemic in any language, there are no unitary symbols for them and linguists have less training in discerning them from other segments. Hearing unreleased stops in continuous speech, perhaps intervocalically, is quite hard. Other kinds of variation, especially variation in consonant manner, is also quite hard to hear. The next section consists of a more detailed discussion of /k/ to set the stage for examining its variation in spontaneous American English speech.

1.2 American English /k/

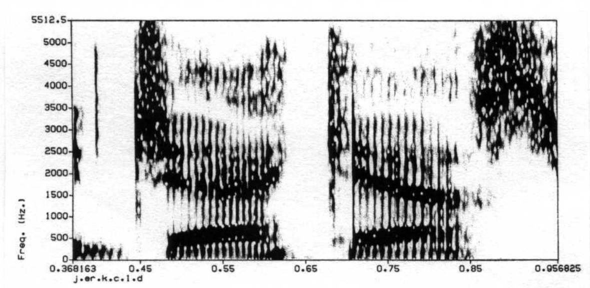
American English /k/ displays more variation than is typically reported. As a velar stop, /k/ should be articulated with a full closure, formed by raising the tongue dorsum to the velum (soft palate). The midsagittal section of the vocal tract below in (1) illustrates the position of the articulators for a canonical stop /k/ and the spectrogram in (2) illustrates the pattern resulting from articulation of a canonical /k/. The /k/ consists of a silent closure, release burst and brief period of aspiration.

(1) Midsagittal section of /k/ articulation (*Language Files* 1999)



(2) Spectrogram of canonical /k/ in the word *jackass*

closure ↓ ↓ release burst



d ʒ æ k æ s

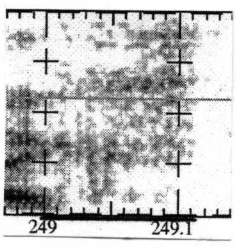
Although /k/ is supposed to be a complete stop, it is frequently realized with other manners of articulation, especially in spontaneous speech. This is variation above and beyond the commonly-noted /k/ variation of aspiration or fronting. According to Ohala (1996:206): "The more we look at connected speech in detail, the larger the 'zoo' of strange and exotic phonetic animals becomes." Ohala is certainly correct in that some of the most interesting, and perhaps even shocking, realizations of /k/ are found in spontaneous speech, where /k/ is not necessarily a complete stop. The two main ways in which /k/ could be less than a stop is for it to lose its closure and to lose its burst. An especially broad range of realizations of /k/ comes from MIT's American English Map Task recordings.

1.3 MIT American English Map Task

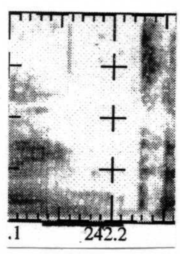
The MIT Map Task recordings, based on the Human Communication Resource Centre (HCRC), or Edinburgh, Map Task (e.g. Anderson et al. 1993) were organized by Olga Goubanova, when she was a visitor at MIT in the Speech Communication Group of the Research Laboratory of Electronics in 1999. She recruited eight close female friends in their late teens and early twenties to participate. Seven of the speakers were from the northeastern United States and one was from California. The Map Task is done in pairs of speakers, with one direction giver and one direction follower. Each member of the pair is given a slightly different version of a map of the same fictional place. The direction giver's map has a route marked which she must convey to the direction follower, who will reproduce it on her map. Several complications make this task quite difficult. The maps do not show precisely the same set of landmarks and the landmarks sometimes have slightly different names (e.g. fast-running river vs. fast-flowing stream) on each map. The complications are intended to encourage dialogue between the speakers, and prevent a simple monologue from the direction giver. In the MIT recordings, the participants had no eye contact. The dialogues were recorded onto digital tape at a sampling rate of 16 kHz, with both dual channel and single channel versions eventually stored. Manipulations of the files were done using Waves+/ESPS signal analysis software. In the winter of 2001, I orthographically transcribed the sixteen Map Task conversations, which totaled two and a half hours of running speech.

In the MIT Map Task, /k/ shows a great deal of variation in its realization. Some of this range of variation is illustrated in the spectrograms in (3) through (6). In (3), /k/ is an incompletely closed stop with some frication noise in *it's like three inches*; in (4), /k/ is almost fully voiced in *like diagonally*; in (5), /k/ is realized as an approximant, with full voicing and just an approximation of closure in *like*; in (6), /k/ is realized as a voiced non-velar fricative in *kinda*. In (3) through (5), the second and third formants of the preceding vowel come together, the velar pinch characteristic of a velar constriction, in the transition to the consonant. In (6), the formants show no evidence of a velar place of articulation; rather, the formants interpolate almost directly between the surrounding vowels, a pattern more characteristic of a glottal fricative.

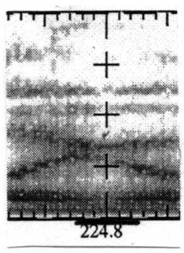
(3) /k/ as incomplete stop with frication ↓



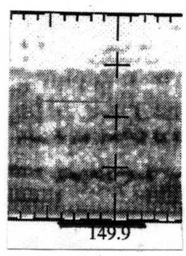
(4) /k/ with voicing ↓



(5) /k/ as an approximant ↓



(6) /k/ as voiced glottal fricative ↓



The spectrograms of /k/ from the Map Task show that /k/ can be realized with several kinds of differences from the canonical stop, either with an incomplete closure, as voiced, as an approximant, or as a non-velar fricative. All of these realizations are at odds with the fact that the only widely acknowledged allophones of /k/ are the fronted and unreleased versions, but have little to do with manner of articulation. While the conditioning environments for the precise manners of articulation of /k/ are not yet clear, it is clear is that the variables influencing the realizations are quite complex.

This work on /k/ is not the first to demonstrate variation from citation form. Other researchers have shown that consonants vary from their citation forms in numerous ways. These include differences in degree of constriction (e.g. Butcher 1996, Crystal & House 1988, Engstrand & Lacerda 1996, Helgason 1996), presence or absence of a burst (e.g. Duez 1995), aspiration, frication, affrication noise, spectral center of gravity (van Son & Pols 1999), amount of formant structure visible, and vocal fold vibration. Although consonants can be produced with other than their canonical forms, they are not necessarily perceived any differently. Whalen, Best & Irwin (1997) found that allophones of /p/ are stored as members of a single underlying category. Even in realizations that are not consistent with the citation forms, enough cues must remain to allow speakers access to the underlying segment.

1.4 Acoustic study of variation in English and Spanish

To study the variation in consonants from a controlled corpus, an acoustic study of variation in 20 American English and 17 Mexican Spanish consonants (Lavoie 2001) was carried out. This study reveals the extent of consonant variation even in relatively careful read laboratory speech. Consonants were all intervocalic in disyllabic words. The position in word (initial or medial) and the lexical stress (either onset to the syllable with primary lexical stress or not) were manipulated. Examples of words studied for /k/ in English and Spanish appear below, with the syllable bearing primary lexical stress underlined:

(7)	**Words for English /k/**		**Words for Spanish /k/**	
	Stressed	**Unstressed**	**Stressed**	**Unstressed**
Initial	*Cocoa*	*Cocaine*	*Coco* coconut	*coCól* bread roll
Medial	*maCaque*	*jaCKass*	*taCón* heel of shoe	*taCo* heel, cue

Five native English speakers and four native Spanish speakers read four repetitions at a time of the words embedded in a carrier phrase. With regard to reduction, the recording paradigm contains aspects that favor reduction and aspects that do not. The most important factor working against reduction is the fact that speakers read the words into a microphone in a soundproof recording booth in the Cornell Phonetics Laboratory. Factors favoring reduction are the fact that most of the speakers were friends of the experimenter, bringing a degree of ease to the recording task. Speakers were also quite relaxed and comfortable with the task because they had read the lists of words in carrier phrases several times already and the words were highly predictable, another conditioning factor for reduction (see, e.g. Bell et al. 1999).

The recordings were made on analog cassette and digitized onto Sun SparcStations for analysis with Waves+/ESPS software. Six repetitions of each word per speaker were analyzed. In this data, many consonants vary from their citation manner of articulation and degree of constriction. The variation is not categorical variation, but rather continuous or stochastic variation as Pierrehumbert (1994) describes. In fact, even the flapping of American

English /t/ was not categorical in the data, as /t/ in the flapping environment, here medial and unstressed, was not consistently a flap.

1.5 Results

Some consonants in the data were very stable in the sense of being true to their manner of articulation in citation form. These were the voiceless sibilants in both English and Spanish, and the voiceless stops in Spanish. For other segments, though, over 10% of the tokens were realized with an unexpected manner of articulation. The manner is unexpected in the sense that it is not the citation or underlying manner of articulation. These segments are English /k, d, g, dʒ, θ, v, ð, z, ʒ/ and Spanish /tʃ, f, x, ɾ, β, ð, ɣ, j/, summarized and given with example words below in (8). Recall that all segments were examined in four different words, such as shown in (7), not just the one word listed below.

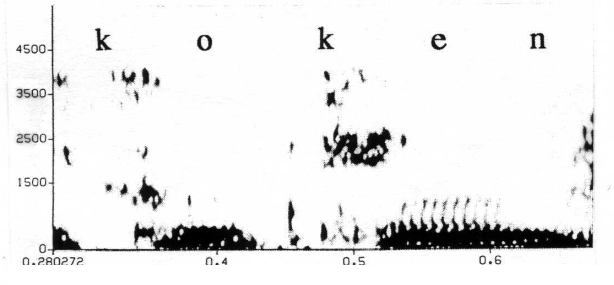
(8) Consonants with more than 10% realized in another manner

English		**Spanish**		
	example word		**example word**	
k	**c**ocaine	tʃ	**ch**ocar	'to strike'
d	**d**isease	f	**f**ocal	'focal'
g	**g**azelle	x	**j**oyón	'large jewel'
dʒ	**J**uly	ɾ	mi**r**o	'I look'
θ	**Th**oreau	β	**b**orrar	'to cross out'
v	**v**ignette	ð	**d**olar	'to chop'
ð	**th**erein	ɣ	**g**ozar	'to enjoy'
z	**Z**aire	j	**ll**orón	'mourner'
ʒ	Bei**j**ing			

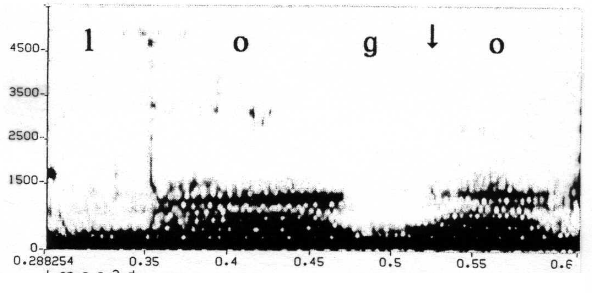
One might well ask, then, if so many segments are not realized in their appropriate citation form, what kind of variation do they exhibit?

Stops in the data are not always complete. They may be "stopless," that is, lack complete seals (e.g., Crystal & House 1988, Shockey & Gibbon 1993) or they may lack release bursts. Note that although numerous English voiceless stops were not complete stops and exhibited noise when there should have been silent closure, they are not true fricatives. Shockey & Gibbon (1993) show that the oral airflow in incompletely closed stops is much less than the airflow for a true fricative. English voiced stops were often realized as approximants, lacking noise but appearing more like glides. The following spectrograms illustrate some of the variation in stops.

(9) English /k/ with incomplete seals and frication noise (first /k/ especially)



(10) English /g/ without release burst



Phonemic fricatives are not always true noisy fricatives; they do not always have concentrated noise in the frequency range expected based on the place of articulation. Since the fricative manner falls between complete closure and approximation of closure, fricatives have two obvious directions in which to vary: they could become more closed or more open. English voiceless fricatives /f, θ/ were sometimes produced with greater closure, as stops. Some English and Spanish fricatives were produced with less closure, as approximants or glides rather than noisy fricatives. In English, the non-sibilant voiced fricatives /v, ð/ show the most variation in their manner of realization by varying in either direction, such that there were realizations as stops, true fricatives, approximants, or glides. Some of the variation in manner of articulation clusters in particular prosodic positions.

1.6 Patterning in terms of sonority

Some of the consonant variation can be explained with the phonological concept of sonority. Sonority highlights the tendencies seen in the data, but it is not entirely descriptive of the data, especially in the case of /k/ which shows variation regardless of the environment. With many segments of both English and Spanish, prosodic position tends to condition differences in sonority of the elements realized, in the following way:

- in unstressed or medial positions:
  consonants may receive *more sonorous realizations*

- in stressed or initial positions:
  consonants tend toward *less sonorous realizations*

Although sonority is a mature concept (e.g. Jespersen 1904), researchers still differ in the level of detail to include in a sonority hierarchy. The minimal sonority hierarchy shown below comes from Zec (1995). Her hierarchy focuses on the important distinction between obstruent, sonorant, and vowel.

(11) Sonority hierarchy (Zec 1995)

obstruent	least sonorous (least open vocal tract)
sonorant	
vowel	most sonorous (most open vocal tract)

Using Zec's hierarchy, it is easy to illustrate how fricatives (a type of obstruent) may be realized, depending on their prosodic position. If the segment is a fricative in citation form, it may receive a less sonorous manifestation when it is word initial or stressed, so it could be a stop (least sonorous type of obstruent). If that same segment is word medial or unstressed, it may receive a more sonorous realization, so it could be an approximant or a glide (sonorant). While these generalizations are useful and descriptive of the results, they do not predict all of the possible consonant variation. Because sonority does not entirely explain all of the consonant variation, some of the other factors influencing the range of variation must be determined. The next section addresses the role of the phoneme inventory in limiting variation.

2 Role of the phonemic inventory in constraining variation

No single factor explains all segmental variation, so it is important to examine numerous factors (see, e.g. Lavoie (to appear)). A variety of factors have already been shown or proposed to influence degree of reduction in words. For this paper, unless otherwise indicated, reduction refers roughly to a segment being realized with less oral closure than in its citation form.

Factors which influence more reduction in a word include:

- optional post-lexical phonology (Byrd 1994)
- predictability: words are shorter if they are more predictable (e.g. Bell et al. 1999)
- collocations: words in common collocations are more reduced (Jurafsky et al. 1998)
- previous use of the word in the discourse: words are shorter on their second utterance in a discourse (Fowler and Housum 1987)
- sex: male speakers reduce more than female speakers (Byrd 1994) and male speakers in TIMIT spoke 6.2% faster than female (Byrd 1994, Jurafsky et al. 1998).

Factors which tend to ensure that words are less reduced, that is, realized very close to their citation forms, or perhaps even hyperarticulated, include:

- disfluency: words are reduced less around a disfluency (Fox, Tree and Clark 1997, Bell et al. 1999)
- first use of a word in a discourse: the first instance of a word in a discourse is less reduced (Fowler and Housum 1987)
- prominence: pitch-accented words are hyperarticulated (de Jong 1995)
- adjacency to higher-level prosodic boundaries: segments that are adjacent to prosodic boundaries of progressively larger size constituents show greater lengthening, indicating that they seldom reduce (e.g., Byrd 2000, Byrd et al. 2000, Byrd & Saltzman 1998).

Some other factors induce differences in reduction, but it is not clear in precisely which direction. These include segmental context and prosodic position. Final consonants reduce less when the following word is vowel-initial (Bell et al. 1999). Glottal consonants /h, /ʔ/ are more vocalic in the vicinity of a pitch accent and more consonantal at phrase boundaries. Gestures have greater magnitude when they are in more prosodically prominent words (Pierrehumbert & Talkin 1992).

The next section examines the role of the language's phoneme inventory, or set of contrastive segments, in constraining phonetic variation. These arguments are informed by earlier work suggesting that the phoneme inventory constrains coarticulation in vowels (Manuel 1990) and constrains variation in consonant place of articulation (Jongman, Blumstein & Lahiri 1985). I will argue that variation in consonant manner of articulation is similarly constrained.

2.1 Vowels

In her work on coarticulation in Bantu languages, Manuel (1990) compared several languages, and found that languages with more crowded vowel systems allow less anticipatory coarticulation than those with less crowded systems. To illustrate, consider an underlying form of /ɑCi/, with C being any consonant. In this case, anticipatory coarticulation between the vowels means that the low vowel /ɑ/ is produced higher than normal, in anticipation of the coming high vowel /i/.

less crowded system → *more* raising of /ɑ/ in anticipation of following high vowel
more crowded system → *less* raising of /ɑ/ in anticipation of following high vowel

These differences in amount of coarticulation are due to the need to keep contrastive phones separate. Manuel (1990:1286) states: "Languages will tend to tolerate less coarticulation just where extensive coarticulation would lead to confusion of contrastive phones." In (12) is an exaggerated diagram of the Ndebele and Sotho vowel spaces, derived from Manuel's results. The circles represent hypothetical extents of coarticulation of the vowel /ɑ/. With only five contrasting vowels, and just one set of mid vowels, Ndebele tolerates raising of /ɑ/. Although the vowel approaches the region of the vowel space occupied by /e, o/, it does not specifically enter the space. Sotho, with its lower mid vowels, on the other hand, cannot tolerate the same amount of /ɑ/ raising. The same amount of coarticulation in Sotho would land /ɑ/ squarely in the space of the lower mid vowels, causing potential loss of contrast.

(12) **Ndebele** **Sotho**

Manuel (1990:1296) summarizes her results:

> […] *the vowel /a/ is more susceptible to anticipatory coarticulation with a following transconsonantal vowel in Shona (LC) and Ndebele (LC), which have no near phonemic neighbors to /a/, than in Sotho (MC), which does have relatively near neighboring and contrasting phonemic vowels. This result is consistent with the idea that coarticulation is limited by output constraints on phones, and that these output constraints are determined, in part, by the need to maintain phonological distinctions in a language.*

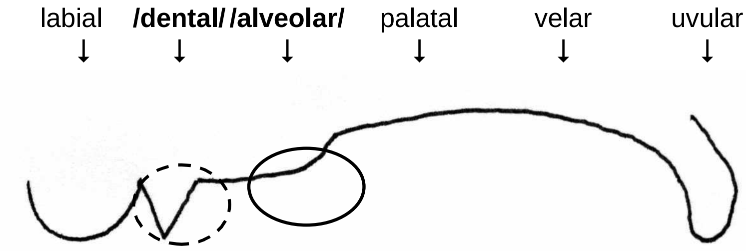
2.2 Consonant place of articulation

Jongman, Blumstein & Lahiri (1985) present the case of dental and alveolar stops in Malayalam, Dutch and English. They demonstrate that dental and alveolar stops in Malayalam can be accurately classified based on stop burst amplitude. They go on to study English alveolar and Dutch dental stops, finding these languages have more variation in stop place than Malayalam, based on stop burst amplitude. They attribute this fact to the three languages' different contrastive uses of the articulatory space for stops, as illustrated below.

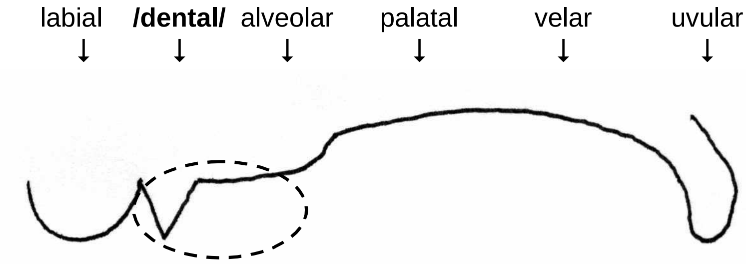
	Contrasting Phonemes	**Realization**
• Malayalam	dental *and* alveolar	distinct dental & alveolar
• Dutch	dental only	dentals range over alveolars
• English	alveolar only	alveolars range over dentals

English and Dutch display more variation in stop place categories than Malayalam does because English and Dutch do not need to distinguish between the dentals and alveolars. In the palate traces below (adapted from Ball & Rahilly 1999:54), possible ranges of closure location are represented with dotted ovals for dentals and solid ovals for alveolars.

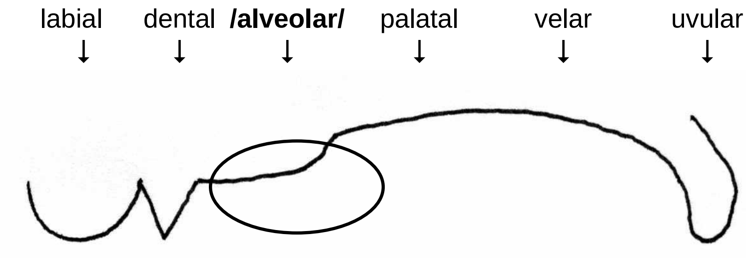
(13) Malayam (distinct dental and alveolar phonemes)



(14) Dutch (just the dental phoneme)



(15) English (just the alveolar phoneme)



2.3 Consonant manner of articulation

Consonant variation has not received much attention with respect to manner. If consonant manner behaves as vowels and consonant place do, the inventory of contrastive elements should constrain manner variation. The voiceless velars in American English and Mexican Spanish provide a perfect test case for this prediction. Examples (16) and (17) below present inventories of voiceless stops and fricatives in Spanish and English. As is clear from the inventories, English has no contrastive voiceless velar fricative, but Spanish does.

(16) Partial phoneme inventory of Spanish

	labial	alveolar	velar
stop	p	t	k
fricative	f	s	x

(17) Partial phoneme inventory of English

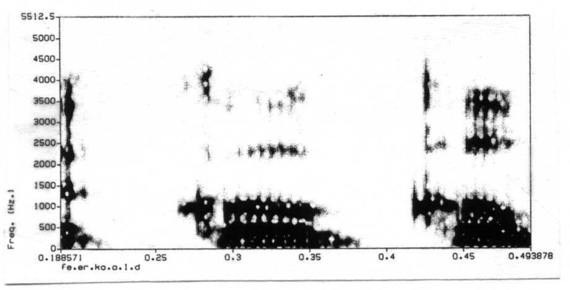
	labial	alveolar	velar
stop	p	t	k
fricative	f	s	

A canonical English /k/ was illustrated earlier in (2) and a canonical Spanish /k/ is illustrated in (18) below. Like the canonical English /k/, Spanish /k/ has a full closure and a release burst. There is no noise leaking through the closure in either of the two instances of /k/ in (18). A canonical Spanish /x/, voiceless velar fricative, is illustrated in (19), with concentrations of frication noise at approximately 1500 and 4000 Hz.

(18) Spanish /k/s in *coca*

closure release closure release

↓ ↓ ↓ ↓

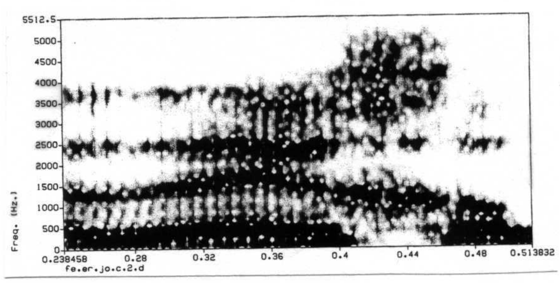


k o k a

(19) Spanish /x/ in *dejo*

frication

↓

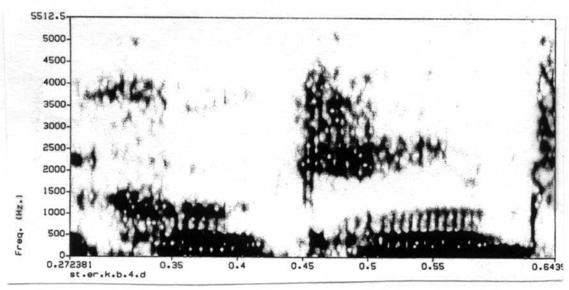


ð e x o

(20) and (21) present examples of English /k/ that differ from the canonical /k/. In (20), /k/ in *cocaine* has frication noise when it should have a silent closure, indicating that the segment is not realized as a complete stop. The realization of /k/ in (21) also shows frication noise, throughout the entire supposed closure. The /k/ realizations given are representative of the data.

(20) English /k/ with frication noise in *cocaine*

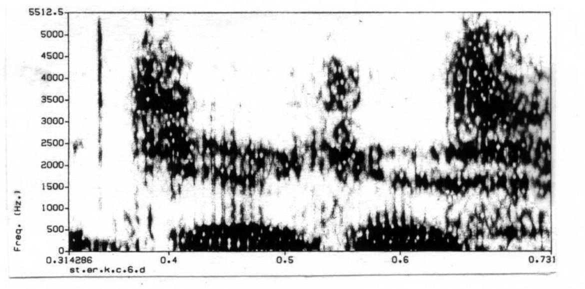
frication closure frication



/k o k e n/
[x o k x e n]

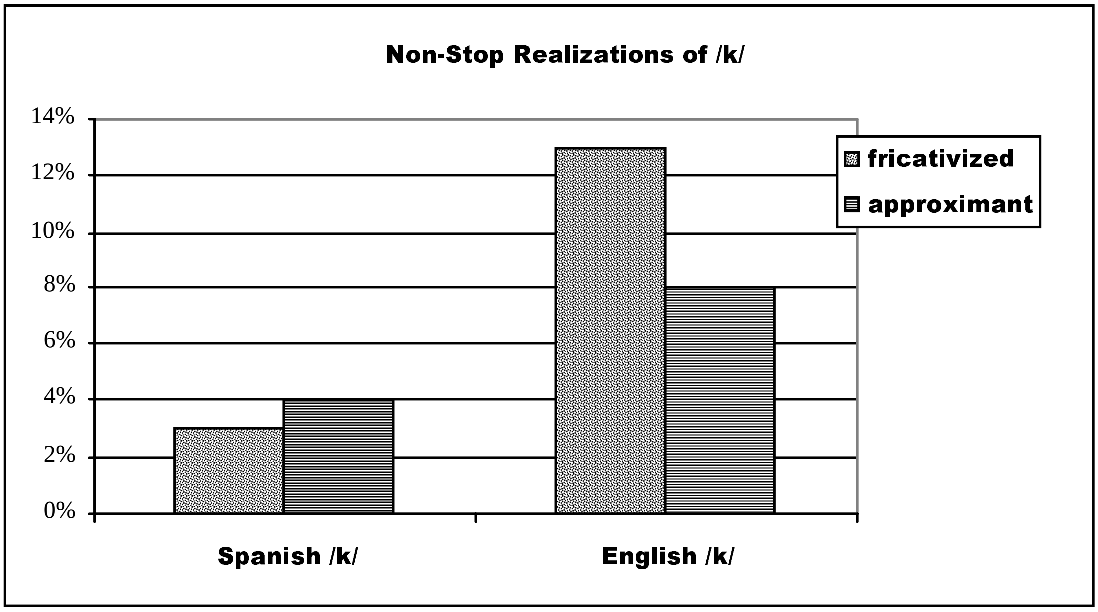
(21) English /k/ with frication in *jackass*

frication



/d ʒ æ k æ s/
[d ʃ æ x æ s]

The chart below summarizes all of the realizations of Spanish and English /k/ in the data that are not canonical stops. In Spanish /k/ is a stop 93% of the time, a fricative 3% of the time and an approximant 4% of the time. In English, /k/ is a stop 79% of the time, some kind of fricative 13% of the time, and an approximant 8% of the time.



These realizations support the hypothesis that, lacking /x/, English /k/ can be realized with noise, while maintaining the percept of a stop. And the presence of the Spanish /x/ prevents /k/ from infringing on /x/'s phonetic space.

Before closing the section on manner of articulation of /k/, slightly more discussion of the incompletely closed or stopless stop segments is in order. Although I use terms like weak fricative, fricative, or fricativized to refer to the stops that show noise leaking through when they should be closed, I do not believe they are true fricatives. The noise usually appears less robust than that for a true fricative. Ken Stevens (p.c., 2001) maintains that the constrictions of true fricatives are optimized to produce sufficiently loud frication, concentrated in the appropriate frequency region. Shockey and Gibbon (1993) examined incompletely closed stops in a corpus of palatometer and airflow data and found that those segments which were not underlyingly fricatives were not realized with as much airflow as those that are truly fricatives. Shockey and Gibbon refer to them as stopless stops. More data are needed to compare real fricatives to these reduced variants of /k/.

3 More predictions about the role of inventory in constraining variation

Having shown the role of the inventory in constraining variation in English and Spanish /k/, I turn to other accessible cases which may be used to test the role of the inventory. These cases include two more in English, the interdental fricatives, and stop releases, as well as predictions for velars in German and French.

3.1 Interdental fricatives

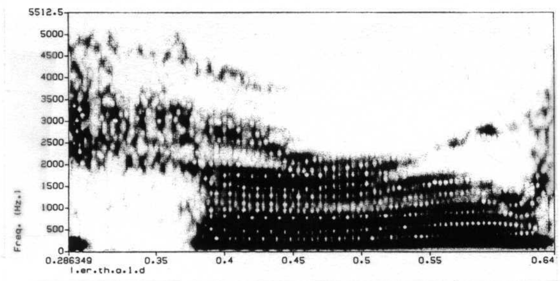
English interdental fricatives are often realized as stops. A common stereotype of Brooklyn, New York speech has the interdental fricatives /θ, ð/ realized as stops, so that *these guys* comes out as *dese guys*. When speaking English, native speakers of languages without interdental fricatives often replace the English interdental fricatives with stops, such that French speakers produce a *thin* that sounds more like *tin*. In its inventory of coronal stops, English has only alveolar stops phonemically, but no contrasting dental stops. Since the English inventory includes phonemic interdental fricatives but no phonemic dental stops, speakers should be free to realize these fricatives as dental stops without risk to the phonemic system. Because of the holes in the English dental category, stop or approximant realizations of the interdental fricatives should be possible.

(22) Partial inventory of English dental and alveolar consonants

	dental	alveolar
stop		t d
fricative	θ ð	s z
approximant		ɾ

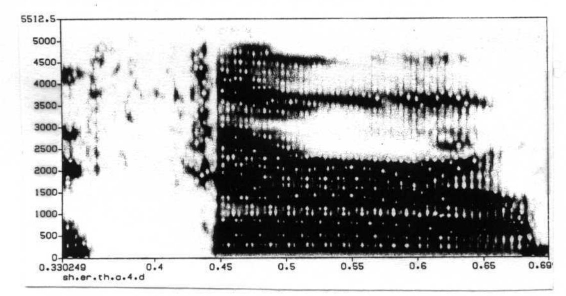
The three examples below all come from the word *thorough* read in the carrier phrase *Please say thorough for me*. (23) has /θ/ as a regular fricative; (24) has an incomplete stop; and (25) has a complete dental stop.

(23) English /θ/ in *thorough* as regular fricative



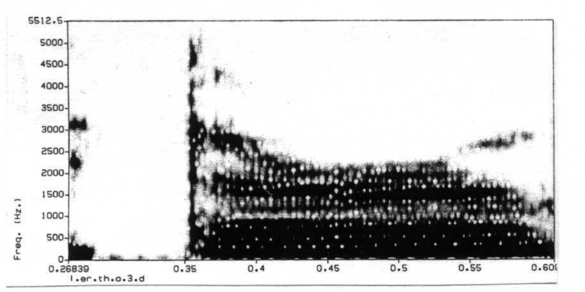
θ ɜ r o

(24) English /θ/ in *thorough* with noise and burst



θ t̪ ɜ r o

(25) English /θ/ in *thorough* as dental stop



t̪ ɜ r o

It is true that the interdental fricative can be realized as a stop. In all three cases, the words sound as if they were produced with a voiceless interdental fricative, with the more stop-like realizations falling below notice. Because the words and the environment are the same in all three cases, either the variation is random or the conditioning factors must be sought elsewhere.

3.2 Stop releases on American National Public Radio (NPR)

Another example in American English where speakers manipulate non-contrastive elements is on National Public Radio (NPR). The style of many NPR announcers shows very robust releases to utterance-final stops. The releases are so robust that the stops seem to be affricated. Based on the hypothesis presented to this point, affricated stops would be fine because, except for the palatoalveolar affricates, there are no contrastive affricated stops. Mark Tiede (p.c., 2001) suggests that it is precisely in the non-contrastive areas that language tends to "bulge out" and speakers show some creativity.

3.3 Predictions for German and French velars

German and French are other possible languages to look at with respect to this hypothesis. The appropriate sections of the phoneme inventories for German and French are given below:

(26) German

	velar
stop	k
fricative	x

(27) French

	velar
stop	k
fricative	

Since German has both a phonemic velar stop and a phonemic velar fricative, the German /k/ should always be realized as a stop to avoid overlap with the fricative phoneme. Most varieties of French, on the other hand, have no phonemic velar fricative and so /k/ could vary from a stop. While a prediction based simply on the inventory of French velars is easy, it is also important to remember that French stops typically have very clear release bursts. To achieve such a salient release, the speaker must make a complete closure, so there is an additional factor requiring the French /k/ to be realized as a full stop. In all cases, it is important to consider the language-specific articulatory tendencies and their possible impact on variation.

This paper has illustrated the wider range of variation that is seen in consonants and shown that the language's phoneme inventory plays a role in constraining that variation. This variation has a number of possible applications.

4 Applications and theoretical implications of consonant variation

The results of studies of consonant variation can be useful in many areas of applied and theoretical linguistics. In linguistics applied to speech technology, a model of variation could yield improved automatic speech recognition and more natural speech synthesis. In second language acquisition, a better understanding of variation may help learners understand colloquial speech, as well as improve their accents and sound more native. In studies of speech style or rate, consonant variation may be able to serve as a quantitative diagnostic of the style or rate, which would be helpful in some areas of sociolinguistics.

Consonant behavior has good potential as a diagnostic of speech style. Up to now, there have been attempts to develop vowel-based diagnostics of style, focusing on vowel reduction and centralization as hallmarks of an informal style. Vowel-based diagnostics are very time-consuming because of the need to calculate formant frequency baselines for each vowel and subject before calculating differences from those baselines. A consonant-based diagnostic would not require the setting of baselines; rather, the citation form of the consonant could be taken as the baseline and differences from that form could be calculated directly and quickly.

Consonants are important not only for applied research, but also to theoretical questions. Theoretically, consonant behavior can have implications for the representation of segmental variation, the relationship between consonant and vowel elements, the correspondence between places of articulation, and the impact of categorical perception on consonants. At the extreme, the results of studies of variation may cause us to rethink our underlying representations. Certainly, it is important to determine where in the phonetics or phonology the range of possible segment variation resides, if it does reside within the phonetics or phonology. The range of possible or acceptable variation is something that might need to be represented.

Phonologically, the study of reduction in consonants can yield insights into the relationship between the consonant and vowel tiers in some versions of autosegmental phonology, in particular the degree of linkage between the tiers. Does reduction in vowels imply concomitant reduction in consonants and/or vice versa? Studies of reduction of both consonants and vowels will help answer the very important question of whether or not consonants and vowels reduce in tandem and if a measure of reduction can be correctly taken to imply reduction in the other. The question of how detailed phonetic place of articulation fits with categorical phonological place of articulation is also important in consonant variation. For example, this study classes the interdental fricatives as dental to oppose them to alveolars. While grouping interdental and dental seems appropriate in this case, there are other groupings that may be more problematic. For example, how do the correspondences between the various other coronal places work? Will the presence of a phonemic palatoalveolar affricate prevent an alveolar stop from affricating? Will the presence of a phonemic labiodental fricative prevent a labial stop from being realized as a labial continuant? Results of studies of consonant variation may yield different natural groupings of places of articulation.

Because consonants are often produced with other than their citation manner of articulation, but listeners are seldom aware of it, categorical perception must be playing a role. Categorical perception of manner of articulation is an area where more detailed perceptual testing is needed. When consonants are produced with other than their expected manner of articulation, what do listeners detect? Do listeners perceive only the intended phoneme or can they judge where the realization fits on a continuum of possible acceptable realizations? Perceptual studies will be important because categorical perception obscures more consonant variation than was previously thought. Determining the range of acceptable realizations for consonants can provide additional data to test the role of phoneme inventory in constraining variation.

5 Conclusions

Consonants show a great deal of variation from their citation forms in connected speech but that phonetic variation respects the phoneme inventory in some ways. A comparison of English and Spanish /k/ has shown the role of inventory in constraining variation. Spanish /k/ may not receive a more open articulation because that would interfere with the contrasting velar fricative /x/. English /k/ is free to receive an open articulation because English has no contrasting velar fricative. Likewise, English interdentals are free to be realized as stops because English has no contrasting dental stops. About 10% of the variation is accounted for by reference to the inventory. Although that is certainly not all the variation, it is significant when considering that many factors affect segmental realization. Studies of consonant variation contribute both to phonetic and phonological theory, offering potentially rich theoretical rewards.

References

Anderson, A.H., M. Bader, E.G. Bard, E. Boyle, G. Doherty, S. Garrod, S. Isard, J. Kowtko, J.M. McAllister, J. Miller, C. Sotillo, H. Thompson, R. Weinert. 1993. The HCRC map task corpus. *Language and Speech* 34, 351-366.

Ball, M. and J. Rahilly. 1999. *Phonetics: The Science of Speech*. London: Arnold Publishing.

Barry, W.J. 1996. Some fundamental problems of looking at connected speech. *Arbeitsberichte* 31. Institut für Phonetik und Digitale Sprachverarbeitung, Kiel University, 113-118.

Bell, A., D. Jurafsky, E. Fosler-Lussier, C. Girand, and D. Gildea. 1999. Forms of English function words—Effects of disfluencies, turn position, age and sex, and predictability. *Proceedings ICPhS 14*, 395-398.

Brown, G. 1990. *Listening to Spoken English*. 2nd ed. London and New York: Longman.

Butcher, A. 1996. Some connected speech phenomena in Australian languages: Universals and idiosyncrasies. *Arbeitsberichte* 31. Institut für Phonetik und Digitale Sprachverarbeitung, Kiel University, 83-104.

Byrd, D. 2000. Articulatory vowel lengthening and coordination at phrasal junctures. *Phonetica* 57, 3-16.

Byrd, D. 1994. Relations of sex and dialect to reduction. *Speech Communication* 15, 39-54.

Byrd, D. A. Kaun, S. Narayanan, and E. Saltzman. 2000. Phrasal signatures in articulation. In M. Broe & J. Pierrehumbert (eds) *Papers in Laboratory Phonology V: Acquisition and the Lexicon*, 70-87. Cambridge: Cambridge University Press.

Byrd, D. and E. Saltzman. 1998. Intragestural dynamics of multiple prosodic boundaries. *Journal of Phonetics* 26, 173-199.
Crystal, T. and A. House. 1988. The duration of American English stop consonants: an overview. *Journal of Phonetics* 16, 285-294.
de Jong, K. 1995. The supraglottal articulation of prominence in English. *JASA* 97, 491-504.
Duez, D. 1995. On spontaneous French speech: Aspects of the reduction and contextual assimilation of voiced stops. *Journal of Phonetics* 23, 407-427.
Engstrand, O. and F. Lacerda. 1996. Lenition of stop consonants in conversational speech: Evidence from Swedish. *Arbeitsberichte* 31. Institut für Phonetik und Digitale Sprachverarbeitung, Kiel University, 31-41.
Fougeron, C. 1998. *Variation articulatoire en début de constituants prosodiques de différents niveaux en français*. Paris III Doctoral Dissertation.
Fougeron, C. and P. Keating. 1997. Articulatory strengthening at edges of prosodic domains, *Journal of the Acoustical Society of America* 101, 3728-3740.
Fowler, C.A. and J. Housum. 1987. Talkers' signaling of "new" and "old" words in speech and listeners' perception and use of the distinction. *Journal of Memory and Language* 26, 489-504.
Fox Tree, J.E. and H.H. Clark. 1997. Pronouncing "the" as "thee" to signal problems in speaking. *Cognition* 62, 151-167.
Helgason, P. 1996. Lenition in German and Icelandic. *Arbeitsberichte* 31. Institut für Phonetik und Digitale Sprachverarbeitung, Kiel University, 219-226.
Jespersen, O. 1904. *Lehrbuch der Phonetik*. Leipzig and Berlin.
Jongman, A., S. Blumstein and A. Lahiri. 1985. Acoustic properties for dental and alveolar stop consonants: A cross-language study. *Journal of Phonetics* 13, 235-251.
Jurafsky, D., A. Bell, E. Fosler-Lussier, C. Girand, and W. Raymond. 1998. Reduction of English function words in Switchboard. *Proceedings of ICSLP-98,* Sydney.
Kohler, K. 1990. Segmental reduction in connected speech in German: Phonological facts and phonetic explanations. In W.J. Hardcastle and A. Marchal (eds). *Speech Production and Speech Modeling*. 6-92. The Netherlands: Kluwer Academic Publishers.
Language Files. 1999. Columbus, Ohio: The Ohio State University Press. 8th edition.
Lavoie, L. to appear. Some influences on the realization of *for* and *four* in American English. *Journal of the International Phonetic Association*.
Lavoie, L. 2001. *Consonant Strength: Phonological Patterns and Phonetic Manifestations*. New York: Garland Publications.
Manuel, S. 1990. The role of contrast in limiting vowel-to-vowel coarticulation in different languages. *Journal of the Acoustical Society of America* 88, 1286-1298.
Ohala, J. 1996. The relation between sound change and connected speech processes. *Arbeitsberichte* 31. Institut für Phonetik und Digitale Sprachverarbeitung, Kiel University, 201-206.
Pierrehumbert, J. 1994. Knowledge of variation. *CLS* 30 (Papers from the 30th Regional Meeting of the Chicago Linguistic Society; Volume 2: The Parasession on Variation in Linguistic Theory), 232-256.
Pierrehumbert, J. and D. Talkin. 1992. Lenition of /h/ and glottal stop. In G.R. Docherty and D.R. Ladd (eds*) Papers in Laboratory Phonology II: Gesture, Segment, Prosody*. 90-127. Cambridge: Cambridge University Press.
Shockey, L. and F. Gibbon. 1993. "Stopless stops" in connected English. *Speech Research Laboratory University of Reading Work in Progress*, 1-7.
van Son, R. and L. Pols. 1999. An acoustic description of consonant reduction. *Speech Communication* 28, 125-140.
Whalen, D., C. Best, and J. Irwin. 1997. Lexical effects in the perception and production of American English /p/ allophones. *Journal of Phonetics* 25, 501-528.
Zec, D. 1995. Sonority constraints on syllable structure. *Phonology* 12, 85-129.

Turkish /h/ Deletion: Evidence for the Interplay of Speech Perception and Phonology

Jeff Mielke
The Ohio State University
mielke@ling.ohio-state.edu

1 Introduction

It has been hypothesized that sounds which are less perceptible are more likely to be altered than more salient sounds, the rationale being that the loss of information resulting from a change in a sound which is difficult to perceive is not as great as the loss resulting from a change in a more salient sound. Kohler (1990) suggested that the tendency to reduce articulatory movements is countered by perceptual and social constraints, finding that fricatives are relatively resistant to reduction in colloquial German. Kohler hypothesized that this is due to the perceptual salience of fricatives, a hypothesis which was supported by the results of a perception experiment by Hura, Lindblom, and Diehl (1992). These studies showed that the relative salience of speech sounds is relevant to explaining phonological behavior. An additional factor is the impact of different acoustic environments on the perceptibility of speech sounds. Steriade (1997) found that voicing contrasts are more common in positions where more cues to voicing are available. The P-map, proposed by Steriade (2001a, b), allows the representation of varying salience of segments in different contexts. Many researchers have posited a relationship between speech perception and phonology. The purpose of this paper is to provide experimental evidence for this relationship, drawing on the case of Turkish /h/ deletion.

The first goal of this paper is to test the hypothesis that perception influences phonology. In general, /h/ is a perceptually weak sound, and it is subject to deletion in many languages. Turkish deletes /h/ only in certain segmental contexts that are not obviously related, so its phonology is fertile testing ground for the hypothesis that less salient sounds are more prone to alteration. If perception influences the selection of environments for deletion, /h/ would be expected to delete in environments where it is less perceptually salient, and to be maintained in environments where it is more salient.

Second, it will be shown that speech perception is influenced by phonology. Although speech perception is claimed to be a factor influencing phonology, the perception of categories is not universal. Rather, some aspects of speech perception are affected by a speaker's language background, as will be shown in this paper. The influence of phonology on perception will be apparent in predictable variance between the performance of speakers with different language backgrounds in the perception experiment.

Teasing apart the effects of phonology from the effects of speech perception is crucial to demonstrating that either of these effects actually exists. A correlation does not entail a bidirectional relationship, so this will be shown separately.

Hume and Johnson (2001) propose a general model of the interplay of external forces and phonology, seen in Figure 1. According to Hume and Johnson, a variety of external factors interact with the cognitive symbolic representation of a language's sound system, with the external factors both influencing and being influenced by the phonology. This is represented in Figure 1 by bi-directional arrows between the cognitive representation (phonology) and the external forces of perception, production, generalization, and conformity. In essence, the present study tests the two-way arrow between perception and phonology.

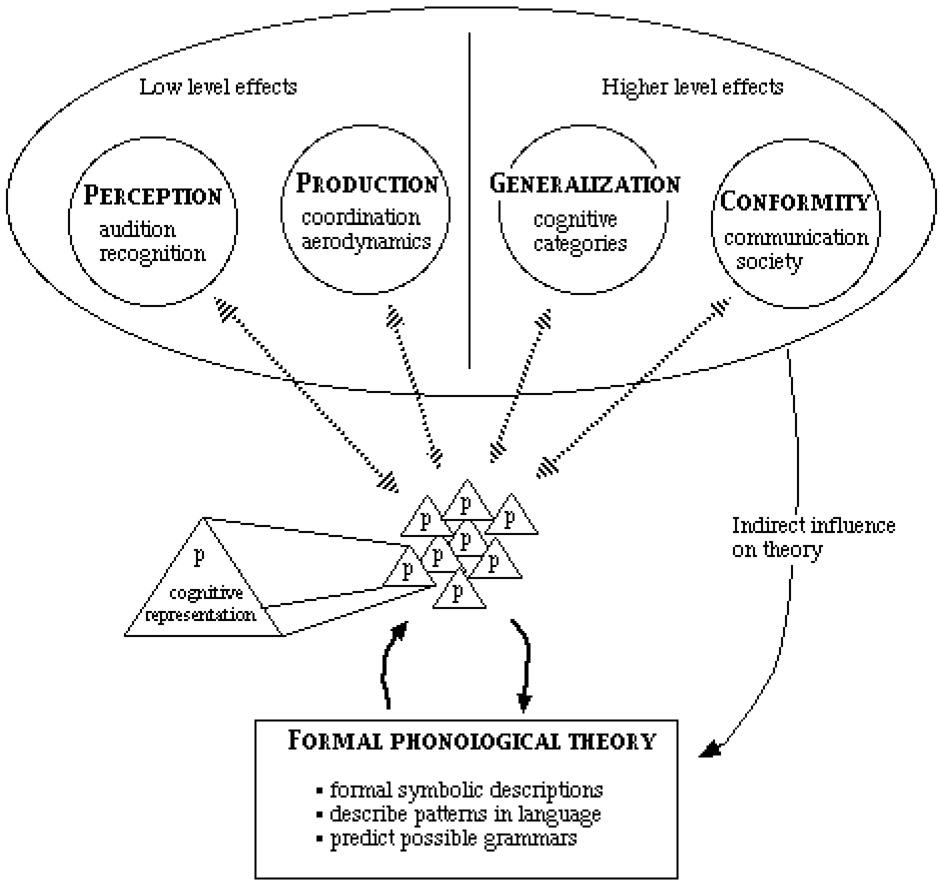


Figure 1. a general model of the interplay of external forces and phonology, broadly defined (Hume & Johnson 2001)

Finally, it will be shown how the perception-phonology relationship can be incorporated into phonological theory. Following Steriade's (1997) use of predicted perceptual salience to generate harmonic constraint rankings based on acoustic cues, the measured perceptibility of /h/ in different environments can be incorporated into a constraint-based account of Turkish /h/ deletion. Assumptions and predictions of such an analysis will be discussed, along with a description of what further empirical data would be useful in order to evaluate the relevance of such an analysis.

2 Turkish /h/ deletion data

/h/ is optionally deleted in fast speech in Turkish, but only in certain segmental contexts (Lewis 1967, Sezer 1986). /h/ is optionally deleted before sonorant consonants (1a), but not after them (1b). When /h/ is deleted from preconsonantal or final position, compensatory lengthening of the preceding vowel occurs, as in (1a). /h/ is optionally deleted after voiceless stops (2b) and affricates (3b), but not before them (2a & 3a). /h/ is optionally deleted before and after voiceless fricatives (4a & 4b), and /h/ is optionally deleted intervocalically (5a), as well as word-finally (5b), but not word-initially (5c).

(1) /h/ is only deleted before sonorants.

a.	fihrist	~	fi:rist	‘index’
	tehlike	~	te:like	‘danger’
	mehmet	~	me:met	proper name
	köhne	~	kö:ne	‘old’
b.	merhum		*merum	‘the late’
	ilham		*ilam	‘inspiration’
	imha		*ima	‘destruction’
	tenha		*tena	‘deserted’

(2) /h/ is only deleted after voiceless stops.

a.	kahpe		*ka:pe	‘harlot’
	sahte		*sa:te	‘counterfeit’
	mahkum		*ma:kum	‘inmate’
b.	ʃüphe	~	ʃüpe	‘suspicion’
	ethem	~	etem	proper name

(3) /h/ is only deleted after voiceless affricates.

a.	ahtʃi		*a:tʃi	‘cook’
b.	metʃhul	~	metʃul	‘unknown’

(4) /h/ is deleted before and after voiceless fricatives.

a.	mahsus	~	ma:sus	‘special to’
	tahsil	~	ta:sil	‘education’
	ahʃab	~	a:ʃab	‘made of brick’
b.	ishal	~	isal	‘diarrhea’
	safha	~	safa	‘step’
	meʃhur	~	meʃur	‘celebrity’

(5) /h/ is deleted intervocalically and word-finally, but not word-initially.

a.	toɦum	~	toum	‘seed’
	müɦendis	~	müendis	‘engineer’
	saɦan	~	saan	‘copper food dish’
	muɦafaza	~	muafaza	‘protection’
b.	timsah	~	timsa:	‘crocodile’
c.	hava		*ava	‘air’

3 Perceptibility hypothesis and predictions

In general, sounds which are less perceptible have been hypothesized to be more likely to be altered than more salient sounds (Hura et al. 1992, Kohler 1990, Steriade 2001). The present study examines the more specific claim that less perceptible sounds are more likely to be deleted than more salient sounds. The focus of this paper is on the behavior of one perceptually weak phoneme in various environments. The hypothesis is that /h/ is less perceptible in environments where it deletes in Turkish than it is in environments where it does not delete. The motivation for loss may be non-perceptual, but perceptibility may determine which sounds are deleted and which sounds are maintained (see Hume & Johnson 2001).

If the hypothesis is on the right track, predictions should be motivated by acoustic and auditory factors that make particular environments perceptually poor environments for /h/. To make independently motivated predictions about the perceptibility of /h/ in different environments, it is necessary to examine the major cues to the presence of [h], the most common allophone of /h/ in Turkish. [h] is marked by aperiodic noise in the F2 region, and less energy in the F1 and F0 regions than would be expected for a sonorant consonant.

In isolation, these cues are very weak, but /h/ is more salient in contexts where it contrasts syntagmatically with surrounding segments. A visual metaphor for the perceptibility of a segment with weak internal cues is given in Figure 2. A white letter is most salient against a dark background that contrasts with it as in (a). It is less salient against a background that contrasts less, as in (b), and least salient against a white background that does not contrast with it at all, as in (c).

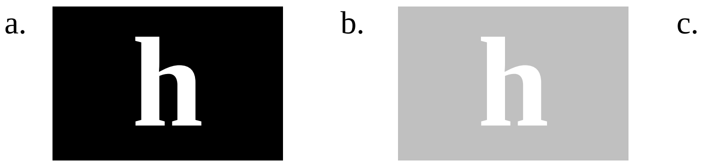


Figure 2. syntagmatic contrast (visual)

Similarly, /h/ is relatively salient in environments where there is more contrast with surrounding segments. /h/ has weak internal cues, but its noise and lack of strong F1 and F0 resonances become more salient in the context of adjacent voiced sonorants lacking aperiodic noise (Figure 3). It is less salient when it is adjacent to segments which also bear aperiodic noise or lack voicing.

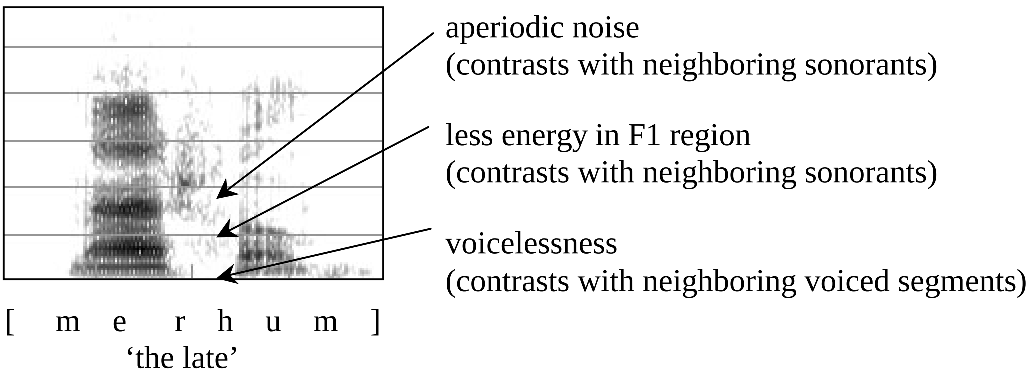


Figure 3. syntagmatic contrast (auditory)

Considering these cues, specific facts about environments lead to predictions of relative salience which are consistent with Turkish deletion patterns. The hypothesis that /h/ is less salient in environments where it deletes is validated if a substantial number of predictions are correct.

One natural prediction is that /h/ should be more salient before a vowel. Auditory nerve fibers exhibit a greater response at the onset of a stimulus signal (such as a vowel) than at the offset (Bladon 1986, Wright 1996). Fujimura et al. (1978) found that CV transitions provide better place cues than VC transitions (see also Ohala 1992). This leads to the prediction that /h/ should be less perceptible before a sonorant consonant than after, because when /h/ follows a sonorant, it is prevocalic. This is consistent with Turkish /h/ deletion patterns.

The fact that the opposite deletion pattern exists for voiceless stops and affricates can be explained on the basis of the fact that /h/ is immediately adjacent to aspiration or frication when it follows a voiceless stop or affricate (Figure 4, left), whereas when /h/ precedes a voiceless stop or affricate, it is separated from the noise by the stop closure (right). This leads to the prediction that /h/ should be less perceptible after these sounds than before them. This is also consistent with Turkish /h/ deletion patterns.

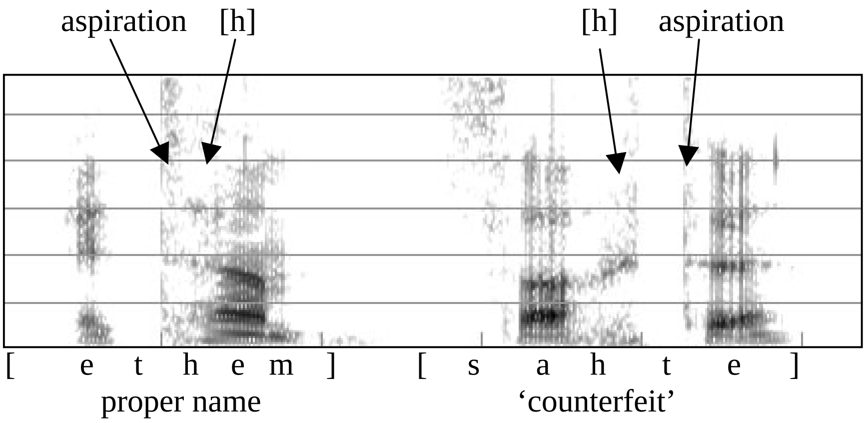


Figure 4. The aperiodic noise of [h] is masked by the aspiration of a preceding voiceless stop (left). Aspiration is hypothesized to be less disruptive when the voiceless stop follows (right).

Another prediction is that /h/ should be more salient after sonorants than after any type of voiceless obstruent, because voiceless stops, affricates, and fricatives all feature noise at the right edge. Intervocalic /h/ is hypothesized to be less salient than initial /h/ for a different reason, namely that intervocalic /h/ is realized as a voiced sound in Turkish, and this reduces the contrast with the environment in a different way, as in Figure 5.

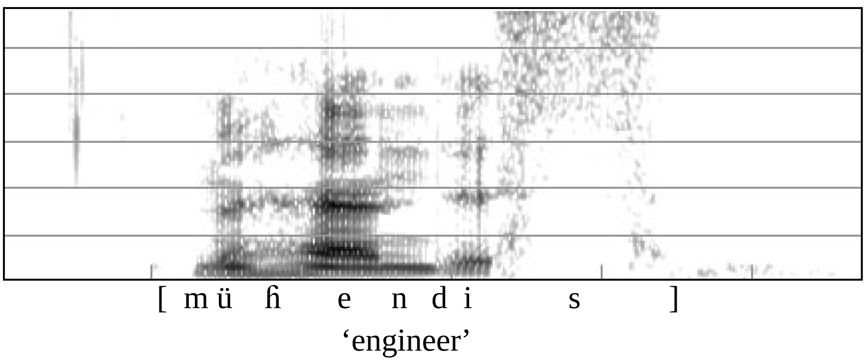


Figure 5. Intervocalic /h/ is hypothesized to be less salient due to voicing.

If these predictions are correct, and /h/ is less salient in the environments where it deletes, then it can be concluded that perception and phonology are related. However, establishing the nature of this relationship is not as simple as proving or disproving the hypothesis. There are at least four logically possible ways for perception and phonology to be related. First, perception and phonology could be completely unrelated, meaning there is no relationship between the ability of speakers to perceive sounds and the way those sounds are used in language. Second, perception could influence phonology, meaning that languages tend to lose contrasts that are not very perceptible. Third, phonology could influence perception, meaning that speakers tend to be less able to discriminate phonetic differences that are not phonologically contrastive in their language. Fourth, phonology and perception could influence each other, meaning that languages tend to lose contrasts that are difficult to perceive, and that speakers tend to lose their ability to discriminate phonologically insignificant differences.

A crosslinguistic study is necessary to examine which of these possibilities is correct. A perception experiment involving speakers of only one language can show correlation between perception and phonology, but a crosslinguistic experiment is necessary to show causation. If perception influences phonology, then the patterns of deletion in languages with /h/ deletion should be consistent with perceptibility even for speakers of languages without /h/ deletion. If phonology influences perception, then speakers of languages with different phonologies should perform differently in a crosslinguistic perception experiment.

A perception experiment was designed to test the relative salience of /h/ in various phonetic environments for speakers of four languages: Turkish, which allows /h/ in many environments, Arabic, which also allows /h/ in many environments, English, which allows /h/ only in prevocalic environments, and French, which has no /h/ sound at all.

4 Methods

4.1 Stimuli

320 nonword stimuli were produced by a male native speaker of Turkish and recorded using a Shure SM10A head-mounted microphone through a Symetrix SX202 dual mic preamp into a Teac V-427C stereo cassette deck. The stimuli were then digitized at 22050 Hz using a Marantz PMD222 portable cassette recorder.

68 stimuli contained intervocalic consonant clusters consisting of /h/ preceded by one of nine different types of consonant (voiceless stop, voiceless affricate, voiceless fricative, voiced stop, voiced affricate, voiced fricative, nasal, liquid, glide). Another 68 stimuli contained intervocalic consonant clusters consisting of /h/ followed by a consonant. 68 foil stimuli contained a single consonant between vowels and no /h/. 24 stimuli contained /h/ in one of three vowel environments (initial, intervocalic, and final), and 12 corresponding foil stimuli contained no /h/. Half of the consonant foil stimuli contained a long vowel before the consonant and all of the word-final foil stimuli contained a long final vowel. This was to simulate the compensatory lengthening that occurs in Turkish when /h/ is deleted from preconsonantal or word-final position. An additional 80 nontarget stimuli without /h/ were also recorded.

4.2 Listeners

The Turkish speaking subjects consisted of six female and 15 male native speakers of Turkish in Columbus, Ohio, aged 19-33. The English speaking subjects consisted of 17 female and ten male Ohio State University undergraduates, all native speakers of American English. The French speaking subjects consisted of one male and twenty-four female native speakers of

French in Paris, France, aged 18-28. The Arabic speaking subjects consisted of two female and ten male native speakers of Arabic in Paris, France, aged 20-36. Of the twelve Arabic speakers, seven were from Morocco, three were from Algeria, one was from Mauritania, and one was from Jordan. Arabic/French bilingualism is not viewed as a problem for the Arabic subjects, because French has no /h/ sound, and a speaker's language background with respect to /h/ should be the same as for a monolingual Arabic speaker (but very different from a monolingual French speaker).

4.3 Procedures

The stimuli were randomized and played to subjects over Sennheiser HD 420 headphones from a laptop computer in a sound booth. As subjects heard each nonword they were presented on a computer screen with all the segments in the word other than /h/, as in Figure 6, and instructed to click on the point in the nonword where they heard /h/, as in Figure 7, or to click on a button representing no /h/ if they heard no /h/ in the word. An "h" appeared on the screen at the point in the nonword where the subject clicked.

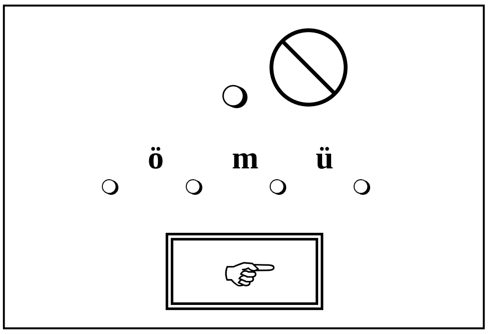


Figure 6. sample screen view: [ömü], [hömü], [öhmü], [ömhü], or [ömüh]?

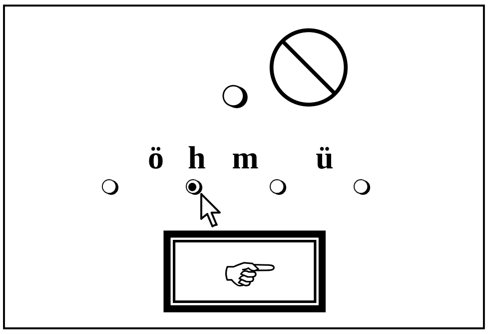


Figure 7. sample response: subject heard [öhmü].

4.4 Data analysis

Sensitivity (d') (Green & Swets 1966, Winer 1971, MacMillan & Creelman 1991) was computed for each subject for each of the 21 environments. d' is a measure of sensitivity based on correct identification and false alarm rates. A d' of zero indicates that correct identification and false alarm rates were the same, that subjects had no sensitivity to the presence or absence of /h/. A positive d' indicates that subjects reported hearing /h/ more often when it was present than when it was not.

5 Results and discussion

The complete results for sensitivity to /h/ in postvocalic environments are given in Figure 8. and the results for prevocalic environments are given in Figure 9.

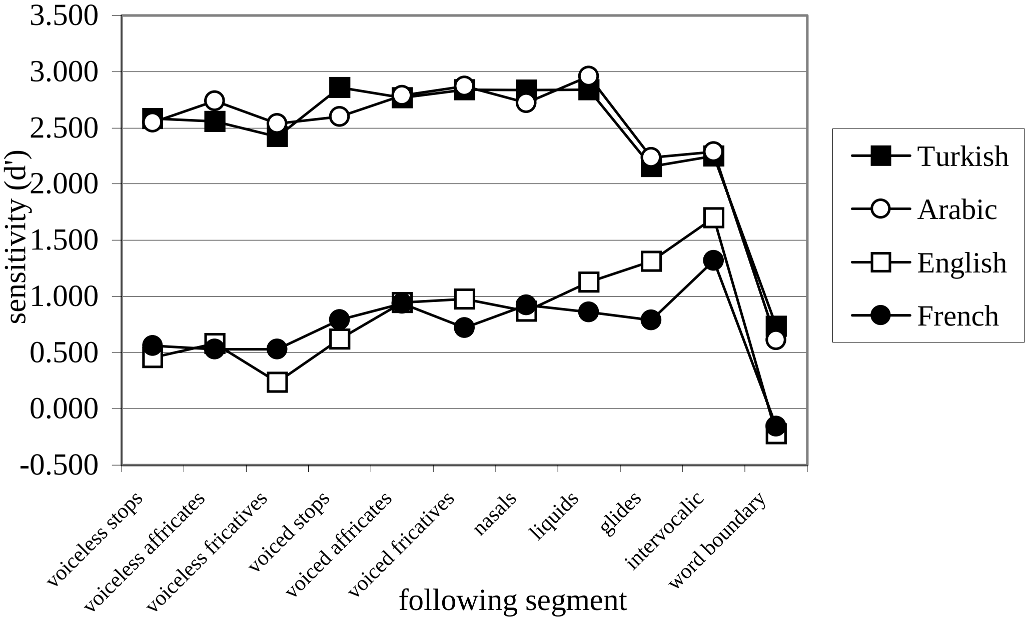


Figure 8. sensitivity (d') to /h/ before context (VhX)

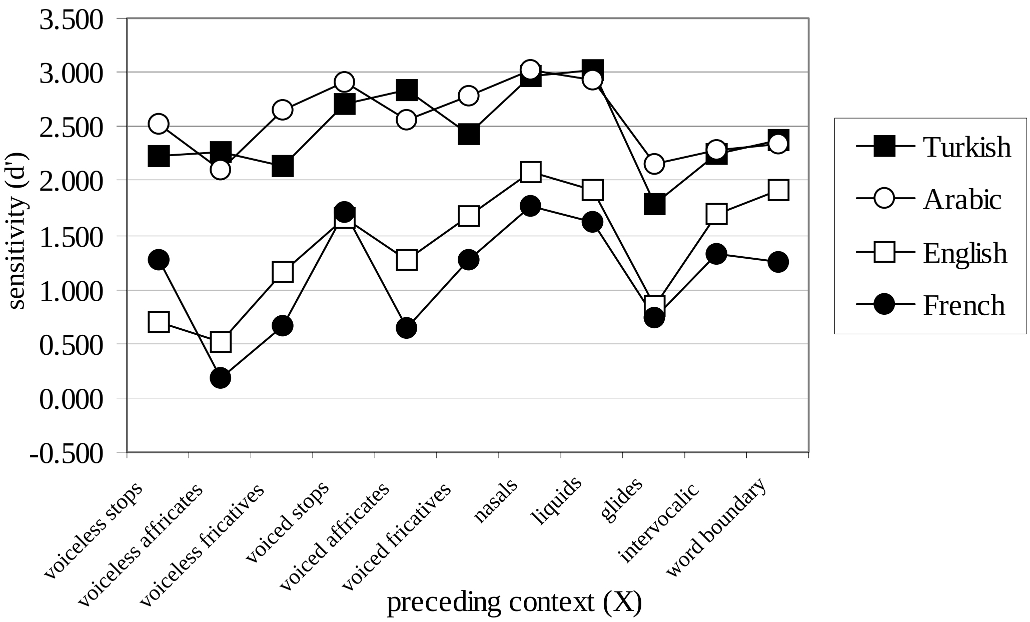


Figure 9. sensitivity (d') to /h/ after context (XhV)

The results are evaluated in terms of the predictions about sensitivity made in the previous section. /h/ was predicted to be less perceptible after voiceless stops and affricates than before them, and less perceptible before sonorant consonants than after them. The results for these environments are displayed in Figure 10.

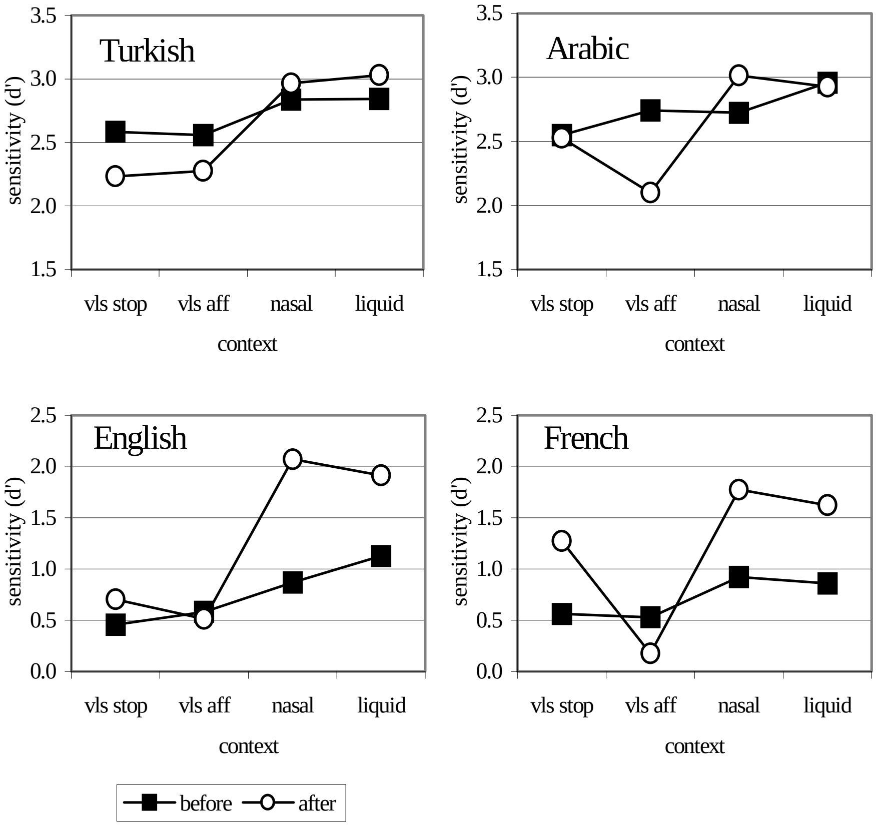


Figure 10. results for voiceless stops, voiceless affricates, nasals, and liquids

A repeated-measures analysis of variance (ANOVA) was performed separately for each language on the subset of data including only the environments before and after voiceless stops, voiceless affricates, nasals, and liquids. Although nasals and liquids were not predicted to differ in their influence on /h/ perceptibility, they were evaluated separately in the experiment. Independent variables were whether or not the consonant was a voiceless stop/affricate or a sonorant (**manner**), and whether the /h/ was preceded or followed by the consonant (**order**). All four groups of subjects showed a main effect for **manner** [Turkish: $F(1,18) = 58.391$, $p < 0.001$; Arabic: $F(1,11) = 13.402$, $p = 0.004$; English: $F(1,20) = 53.352$, $p < 0.001$; French: $F(1,20) = 41.570$, $p < 0.001$], but only English and French listeners showed a main effect for **order** [Turkish: $F(1,18) = .586$; $p = 0.454$; Arabic: $F(1,11) = .448$; $p = 0.517$; English: $F(1,20) = 18.345$; $p < 0.001$; French: $F(1,20) = 10.227$; $p = 0.005$] All four groups of listeners showed a significant interaction for **manner * order** [Turkish: $F(1,18) = 15.090$; $p = 0.001$; Arabic: $F(1,11) = 12.176$; $p = 0.005$; English: $F(1,20) = 15.375$; $p = 0.001$; French: $F(1,20) = 8.054$; $p = 0.010$].

The most interesting aspect of these results is the significant interaction between manner and order. Although /h/ is more perceptible prevocalically for English and French listeners even in cases where it follows a voiceless stop or affricate, the effect of masking by a preceding stop or affricate does, in fact, significantly reduce perceptibility. Order does not have a significant effect overall for Turkish and Arabic listeners because the positive effect of being prevocalic is small enough to be overridden by the masking effect of a preceding stop or affricate.

As is seen in the Turkish results, /h/ is more perceptible after nasals and liquids than before them, but as predicted, the pattern is reversed for voiceless stops and affricates. For each pair of environments shown in figure 10 involving the same type of consonant, deletion occurs in the environment with lowest perceptibility in the pair. The fact that the same general pattern exists for the other three groups of subjects shows that the effect is not specific to Turkish, and therefore that the deletion pattern cannot be solely responsible for the differences in perceptibility. /h/ deletion occurs in environments that Arabic, English, and French listeners also find to be relatively difficult for perceiving /h/, indicating that deletion in Turkish corresponds to a more universal pattern and therefore that perception influences phonology. A number of ways in which the patterns of perceptibility of the other three languages differ from Turkish and from each other show that phonology also influences perception.

All four groups of subjects show significant differences in perceptibility following stops/affricates and sonorants, but perceptibility of /h/ before these consonants is more similar. This is the significant interaction between **manner** and **order**, found in all four groups of subjects. However, for English and French listeners, perceptibility in preconsonantal environments is far lower relative to postconsonantal environments than it is for Turkish and Arabic listeners. This is the main effect for **order**, seen only in English and French listeners, and it shows that for speakers of languages without nonprevocalic /h/, the overriding factor determining the salience of /h/ is whether or not it is followed by a vowel. This is not the case for Turkish and Arabic listeners, and it shows that phonology influences perception. /h/ was predicted to be less salient in non-prevocalic environments, and in the absence of any other factors, this is the case for all four groups of subjects. The difference is overwhelming for the English and French listeners, because the subjects have little or no experience perceiving non-prevocalic /h/.

Additional native language effects also exist. Because the stimuli were produced by a speaker of Turkish, the fact that Turkish nonprevocalic /r/ is pronounced with frication may impede /h/ perception for listeners who are not native speakers of Turkish and do not attribute the frication they hear to /r/. This can be seen in the fact that non-Turkish listeners were marginally less sensitive to /h/ after liquids than after nasals, as opposed to Turkish listeners.

Phonetic differences in the native languages of non-Turkish listeners also played a role. English listeners were less able to detect /h/ after voiceless stops than French listeners. This can be understood by looking at the phonetic realization of phonologically voiceless stops in English and French. French lacks aspirated stops in nonfinal positions (Valdman 1976), and it has been noted in the literature that English voiceless stops are more heavily aspirated than Turkish voiceless stops (Lewis 1967). In a study of noncoronal stop perception, Volaitis and Miller (1992) found that at a fast speech rate comparable to the speech rate of the stimuli for the present experiment, English-speaking subjects recognized labial stops produced with voice onset times up to 87.15 milliseconds and velar stops with VOTs up to 92.10 ms as “normal” voiceless stops, whereas stops with higher VOTs were perceived as “exaggerated”. In the present study, the mean VOT of the Turkish voiceless stop + /h/ sequences in target stimuli was 86 ms, compared with 44 ms for voiceless stops in foil stimuli. The VOTs for both types of stimuli fell within the range Volaitis and Miller found to be perceived as normal for voiceless stops by English speakers. It is not surprising that English listeners had extreme difficulty distinguishing between two types of stimuli that fall into the same perceptual category. That French listeners, who do not have prevocalic aspirated stops, are better able to make this discrimination, is evidence that phonetic differences between native languages contribute to differences in speech perception.

/h/ was predicted to be less perceptible after voiceless stops, affricates, and fricatives, where /h/ deletes, than after sonorant consonants, where /h/ does not delete. The results for these environments are displayed in Figure 11. A second series of ANOVAs was performed on the subset of the data including only the environments after voiceless obstruents and sonorant consonants, with whether or not the /h/ followed a voiceless obstruent as an independent variable. All four groups of subjects showed a main effect for **manner** [Turkish: $F(1,18) = 97.533$; $p < 0.001$; Arabic: $F(1,11) = 18.178$; $p = .001$; English: $F(1,20) = 54.828$; $p < 0.001$; French: $F(1,20) = 37.482$; $p < 0.001$]. Again, it is shown that the environments where /h/ deletes in Turkish are perceptually poor crosslinguistically.

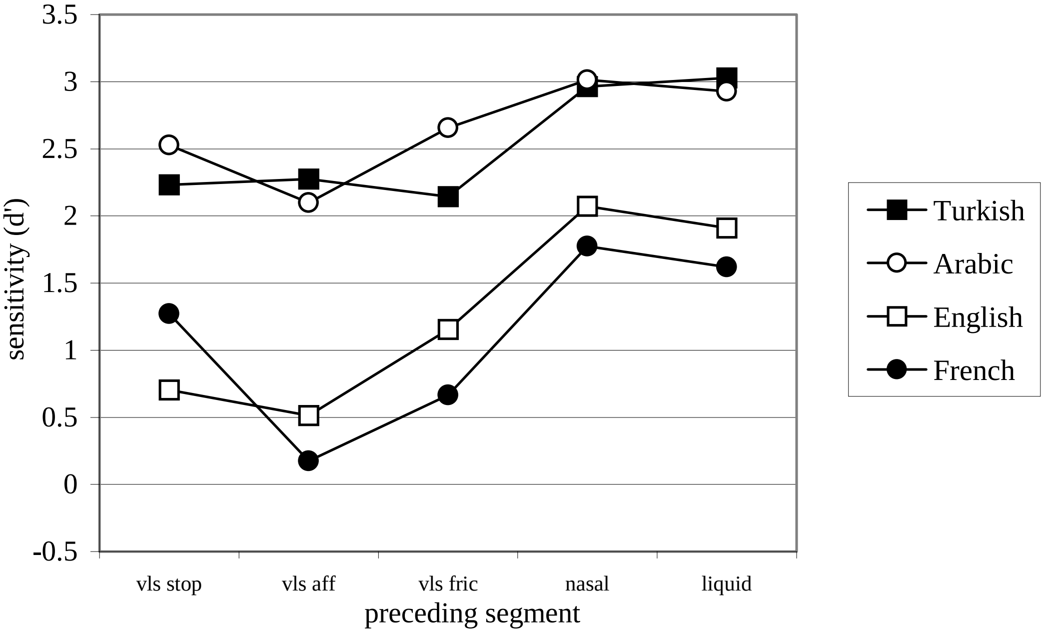


Figure 11. results for after voiceless obstruents and sonorant consonants

In the two of the three environments in Figure 11 where /h/ deletion occurs in Turkish (after voiceless stops, affricates, and fricatives), Arabic listeners were at least marginally more sensitive to /h/ than Turkish listeners, even though the stimuli were produced by a Turkish speaker. This indicates that Turkish speakers may have more difficulty detecting /h/ in environments where it deletes in their language.

Finally, /h/ was predicted to be less perceptible intervocalically, where it deletes, than word-initially, where it does not delete A third series of ANOVAs was performed on the subset of the data including only these two environments, with whether or not the /h/ was intervocalic as an independent variable. None of the four groups of subjects showed a main effect for **intervocalic** [Turkish: $F(1,18) = .656$; $p = 0.429$; Arabic: $F(1,11) = .078$; $p = .786$; English: $F(1,20) = 1.711$; $p = .206$; French: $F(1,20) = .110$; $p = .744$].

It is clear from these results that perceptual salience does not explain why intervocalic /h/ deletes and initial /h/ does not. /h/ is not particularly salient in either environment, but there are reasons other than perceptual salience for not deleting word-initial material. Lexical access is thought to be based on the initial part of a word, and so the left edge of the word is special for word-recognition (Cutler et al. 1985, Marslen-Wilson 1989, Marslen-Wilson & Zwitserlood 1989). Hall (1992) found that there is a tendency for beginnings of words to be particularly robust and less susceptible to phonological processes.

Additionally, further examination of the experimental stimuli found that the phonetic realization of /h/ is more complicated than has been reported in the literature. In addition to being voiced intervocalically, /h/ is also frequently voiced between other sonorant consonants. The impact of this finding on the present study is minimal, however, because contrary to the initial predictions, being voiced does not appear to significantly impact the perceptibility of /h/.

6 Implications for phonological theory

The results of the experiment have demonstrated that there is interplay between perception and phonology. One means of formalizing this relationship and incorporating perception into phonological theory is to use markedness constraints which are aligned to a perceptibility scale (Steriade 1997).

In Steriade's account of [voice] neutralization, constraints prohibiting [voice] contrast in various environments are aligned to a perceptibility scale of voice contrast in those environments. Preserve [voice] is ranked among these constraints, and the result is neutralization of contrast in the environments with the least voicing cues and maintenance of contrast in the environments with the most voicing cues, as shown in (6).

(6) Steriade's (1997) constraint-based perceptual account of [voice] neutralization

A fixed hierarchy of *voice constraints is aligned to a perceptibility scale for voicing.

*voice/ V_[-son] — [voice] is neutralized before obstruents and word-finally.

↓

*voice/ V_#

↓

Preserve [voice]

↓

*voice/ V_[+son] — Contrast is maintained before sonorants.

A similar approach is possible with Turkish /h/ deletion. Constraints prohibiting /h/ in certain environments are aligned to a perceptibility scale of /h/ (for Turkish listeners) in those environments. Max /h/, which prohibits /h/ deletion, is ranked below the constraints prohibiting /h/ in environments where it is deleted and above constraints prohibiting /h/ in environments where it is not deleted, as shown in (7).

(7) Part of a possible constraint-based perceptual account of Turkish /h/ deletion (following Steriade 1997)

A fixed hierarchy of *h constraints is aligned to a perceptibility scale for /h/.

*h/ [vls stop]_V
↓
*h/ V_[nas]
↓
Max /h/
↓
*h/ [nas]_V

/h/ is deleted after voiceless stops and before nasals.

/h/ is maintained after nasals.

Steriade's model predicts that all environments where /h/ deletes would be perceptually poorer than all environments where /h/ is maintained, but this is not the case. /h/ is deleted before liquids [d' = 2.841] and nasals [d' = 2.838] but maintained in two environments where it is less perceptible: before voiceless stops [d' = 2.583] and before voiceless affricates [d' = 2.558]. Steriade's perceptibility scale for voice contrast is based on hypothetical cues rather than experimental results, and in this case, the experimental results show that additional factors are at play. If a hierarchy of markedness constraints were aligned to the perceptibility scale in (8), there would be no place to insert Max /h/ to separate the deletion environments from the non-deletion environments. Considering only perceptibility as a factor, it would be surprising that /h/ deletes before liquids and nasals and that it does not delete word-initially, but this is not as surprising when some psycholinguistic factors are taken into consideration.

(8) Perceptibility scale for Turkish listeners

7 Segment	**8 d'**	**9 Deletion**
[liquid]_V	3.028	no
[nasal]_V	2.964	no
V_[liquid]	2.841	YES
V_[nasal]	2.838	YES
V_[vls stop]	2.583	no
V_[vls affricate]	2.558	no
V_[vls fricative]	2.423	YES
#_V	2.376	no
[vls affricate]_V	2.274	YES
V_V	2.248	YES
[vls stop]_V	2.233	YES
V_[glide]	2.155	YES
[vls fricative]_V	2.144	YES
[glide]_V	1.777	YES
V_#	0.734	YES

In /h/ deletion, as in other cases, the demand for ease of production is opposed by the need to minimize the impact on the information content of an utterance. Perceptual salience is an important factor contributing to the potential for loss of information content, but as mentioned above with regard to initial /h/, it is not the only factor. The special status of initial segments is relevant, and so is the structure of the lexicon and how deletion will impact the distinction between lexical items. Both of these factors are useful in explaining the mismatch between the perceptibility scale and the deletion environments.

For all of the consonants involved in Turkish /h/ deletion, /h/ is deleted either before or after the consonant, whichever is worse perceptually. /h/ can be deleted on either side of the same consonant only when it is perceptually very weak in both positions. From a standpoint of maintaining contrast between lexical items, deletion before or after a consonant for which deletion is already permitted when the consonants are in the reverse order is more costly than the first deletion, i.e., it is more costly to delete /h/ before a voiceless stop if /h/ is already allowed to delete after a voiceless stop, because the result of both deletions is the same: an intervocalic stop.

As an illustration, suppose a hypothetical language that allows /h/ before and after consonants. If, for example, /h/ deletion is prohibited before and after /s/, then /sh/ and /hs/ clusters are in contrastive distribution with each other as well as with a single consonant, /s/. There is a three-way contrast (/sh/ vs. /s/ vs. /hs/). If /h/ deletion becomes possible in one environment, such as /s_/, the distinction between words of the form /VshV/ and /VsV/ is be neutralized. Because /h/ deletion is not permissible in the environment /_s/, a word of the form /VhsV/ would still be distinct from the other two. Now suppose that /h/ deletion becomes permissible in the environment /_s/ as well. The result is that the distinction between /VhsV/ and /VsV/ is lost, as well as the distinction between /VhsV/ and /VshV/. The first deletion led to the neutralization of one contrast (/sh/ vs. /s/), but the second deletion led to the neutralization of two contrasts (/hs/ vs. /s/ and /hs/ vs. /sh/). The second deletion involving the same consonant in the context is more costly in terms of contrast between lexical items.

Considering these additional factors, it is not surprising that in Turkish, /h/ deletion is allowed before liquids and nasals, because although /h/ is quite salient in these environments, /h/ deletion does not severely impact the contrast between lexical items, because /h/ deletion is not allowed after liquids and nasals. Similarly, /h/ deletion is not allowed before voiceless stops and affricates due to the fact that it is allowed after voiceless stops and affricates, and further deletion would impact the contrast between lexical items more severely. Deletion is permitted before and after glides and voiceless fricatives because salience is so low in these environments that the relatively imperceptible difference between the two forms reduces the cost of neutralizing the contrast. In (9), the deletion environments are ranked by a combination of psycholinguistic and perceptual factors: the impact on lexical access due to deleting an initial segment, the impact on lexical contrast due to deleting /h/ when the output of /h/ deletion would be the same as the result of /h/ deletion in an environment where /h/ is less perceptible (a relationship indicated by arrows in the table), and the perceptibility of /h/ in the environment. Within the ranking, there is a certain threshold of cost below which /h/ deletion is permitted.

(9) /h/ deletion environments sorted by impact on lexical access and lexical contrast, and perceptual salience (d’).

10 11 Context	12 Lexical Access	13 Lexical Contrast	14 d’	15 Deletion	16
#_V	✓		2.376	no	
[liquid]_V		✓	3.028	no	
[nasal]_V		✓	2.964	no	
V_[vls stop]		✓	2.583	no	
V_[vls affricate]		✓	2.558	no	
V_[vls fricative]		✓	2.423	YES	DELETION
V_[glide]		✓	2.155	YES	THRESHOLD
V_[liquid]			2.841	YES	
V_[nasal]			2.838	YES	
[vls affricate]_V			2.274	YES	
V_V			2.248	YES	
[vls stop]_V			2.233	YES	
[vls fricative]_V			2.144	YES	
[glide]_V			1.777	YES	
V_#			0.734	YES	

The tableau in (10) shows how this would work in an OT grammar for Turkish fast speech.

Combining psycholinguistic and perceptual factors rather than using a strictly perceptually motivated hierarchy allows a ranking in which the constraints prohibiting /h/ in the environments where it deletes outrank all of the constraints that prohibit /h/ in the environments where it does not delete. Max /h/ can then be ranked at the deletion threshold. This account predicts that other languages with /h/ deletion would have similar scales of markedness constraints but perhaps different deletion thresholds, and Max /h/ would therefore be ranked differently with respect to the other constraints.

Following Nagy and Reynolds (1997), variation can be modeled by allowing a faithfulness constraint such as Max /h/ to be ranked relatively low in fast speech but “float” above some of the higher-ranked markedness constraints for careful speech (with less /h/ deletion). Whether such constraint floating actually occurs is a question that can be answered by further examination of Turkish /h/ deletion in various styles of speech. Positing a floating constraint makes the specific prediction that /h/ deletion would decrease in less casual speech styles by decreasing the number of environments where it occurs (as Max /h/ floats up past the relevant *h constraints), and it is not at all clear that this is what actually happens.

(10) Correct derivation of surface forms for Turkish fast speech.

	*h/ V_#	*h/ [glide]_V	*h/ [vls fricative]_V	*h/ [vls stop]_V	*h/ V_V	*h/ [vls affricate]_V	*h/ V_[nasal]	*h/ V_[liquid]	*h/ V_[glide]	*h/ V_[vls fricative]	**Max /h/**	*h/ V_[vls affricate]	*h/ V_[vls stop]	*h/ [nasal]_V	*h/ [liquid]_V	*h/ #_V
/ethem/ proper name																
a. ethem				*!												
☞ b. etem											*					
/kahpe/ ‘harlot’																
☞ a. kahpe													*			
b. ka:pe											*!					
/mehmet/ proper name																
a. mehmet							*!									
☞ b. me:met											*					
/imha/ ‘destruction’																
☞ a. imha														*		
b. ima											*!					
/mahsus/ ‘special to’																
a. mahsus										*!						
☞ b. ma:sus											*					
/metʃhul/ ‘unknown’																
a. metʃhul						*!										
☞ b. metʃul											*					

In truth, the only crucial rankings motivated by the phonological pattern are the domination of Max /h/ by the ten markedness constraints above it and the domination of the other five markedness constraints by Max /h/. The results of the perception experiment and the assumptions about contrast between lexical items suggest other relative rankings that are important for making crosslinguistic predictions, but how the factors interact is unclear. This matter and the question of whether such a formalization of the relationship between perception and phonology makes any sense at all requires further research into the phonological pattern, specifically with respect to speech style variation (Mielke, forthcoming).

Analyses such as this involving constraints aligned to a perceptibility scale have been used to make crosslinguistic predictions (e.g., Steriade 1997, Kochetov 2001), and in doing so it is important to note that while some aspects of speech perception are universal, some are language-specific, as shown in this study. This finding does not preclude the use of perceptibility scales to make crosslinguistic predictions, but when making such predictions, it must be taken into consideration that perceptibility scales may vary from language to language. While perceptibility scales are language-specific rather than universal, they follow from more general principles and language-specific factors (as discussed in section 4). So perceptibility scales should nevertheless be predictable to some extent from these language-specific factors (see e.g., Mielke 2001).

7 Conclusion

The relationship between speech perception and phonology is seen in two aspects of the perception experiment results. Consistent with the claims of Kohler (1990) and Hura *et al.* (1992) that less perceptible sounds are more prone to alteration, Turkish /h/ has been found to delete in environments where it is least perceptible and to be maintained in environments where it most perceptible, showing that speech perception influences phonology. Furthermore, speakers of Turkish, Arabic, English, and French have been found to differ in their ability to perceive /h/ in ways that are consistent with phonological differences, and to perform in ways that are consistent with phonological and phonetic properties of their native languages, showing that native phonology and phonetics also influence speech perception.

The experimental results demonstrate that a bidirectional relationship exists between speech perception and phonology. It is possible to incorporate a perceptibility scale into a constraint-based account of Turkish /h/ deletion, but whether or not such a formalization is faithful to psycholinguistic reality remains to be seen. Further, while the account benefits from the inclusion of psycholinguistic factors, it is unclear precisely how the psycholinguistic and perceptual factors interact. Numerous open questions about the nature of the perception-phonology relationship exist. Whether the domain of speech perception's influence is limited to diachrony (e.g. Newmeyer 2001) or whether the influence of perception is active in the synchronic grammar is an empirical question that remains open. This question could be answered by examining Turkish speakers' production of words with underlying /h/ in various speech styles (Mielke, forthcoming). If speech perception's influence is active in the synchronic grammar, Turkish speakers may start to delete /h/ first in the least salient environments, and proceed up the perceptibility scale as speech rate increases. If speech perception's influence is limited to diachrony, this would be impossible (cf. Steriade 2001a, b). Speech rate increase would cause an increase in frequency of deletion evenly across deletion environments. A production study could reveal whether speakers have access to a perceptibility scale of environments where /h/ occurs, or are simply aware of classes of environments where /h/ deletion is or is not permitted.

Acknowledgements

This research has been supported in part by a pre-dissertation research travel grant from The Ohio State University. This paper builds on previous work by Lena Ovcharova (1999), and has benefited from comments from Elizabeth Hume, Keith Johnson, Mary Beckman, Donca Steriade, students in the Perception in Phonology seminars at OSU, members of the OSU phonetics and phonology discussion group, and audience members at MOT 2000, the 2000 OSU Colloquium Fest, MCWOP 6, CLS 37, and a 2001 OSU Linguistics Speakers Series talk. The French and Arabic experiments were made possible by the help of Nick Clements and Annie Rialland.

References

Bladon, A. 1986. Phonetics for hearers. In G. McGregor (ed.), *Language for Hearers.* Oxford: Pergamon, pp. 1-24.

Cutler, A., J.A. Hawkins & G. Gilligan. 1985. The suffixing preference: a processing explanation. *Linguistics* 23:723-58.

Fujimura, O., M. J. Macchi and L. A. Streeter. 1978. Perception of Stop Consonants with Conflicting Transitional Cues: A Cross-linguistic Study, *Language and Speech*, Vol 21, Part 4.

Green, D. M. and J. A. Swets. 1966. *Signal detection theory and psychophysics.* New York: Wiley.

Hall, C. 1992. Integrating diachronic and processing principles in explaining the suffixing preference. In *Morphology and Mind: A unified approach to explanations in linguistics*. New York: Routledge.

Hume, E. and K. Johnson. 2001. A Model for the Interplay of Speech Perception and Phonology. In E. Hume and K. Johnson (eds.) *The Role of Speech Perception in Phonology*. New York: Academic Press.

Hura, S. L., B. Lindblom & R. Diehl. 1992. On the role of perception in shaping phonological assimilation rules. *Language & Speech* 35:59-72.

Kochetov, A. 2001. Production, perception, and emergent phonotactic patterns: A case of contrastive palatalization. PhD dissertation, Toronto.

Kohler, K. 1990. Segmental reduction in connected speech: Phonological facts and phonetic explanations. In W.J. Hardcastle & A. Marchal (eds.) *Speech Production and Speech Modeling* (pp. 69-92). Dordrecht: Kluwer Academic Publishers.

Lewis, G. L. 1967. *Turkish Grammar.* Clarendon Press. Oxford.

MacMillan, N. A. and D. Creelman. 1991. *Detection Theory: A User's Guide.* Cambridge. New York.

Marslen-Wilson, W.D. 1989. Access and integration: Projecting sound onto meaning. In W.D. Marslen-Wilson (ed.), *Lexical Representation and Process.* Cambridge, MA: MIT Press. 3-24.

Marslen-Wilson, W.D., & P. Zwitserlood. 1989. Accessing Spoken Words: The importance of word onsets. *Journal of Experimental Psychology: Human Perception and Performance.*

Mielke, J., 2001. Perception in phonology: the case of Turkish /h/ deletion, in Andronis, Mary, Christopher Ball, Heidi Elston and Sylvain Neuvel eds. *CLS 37: The Panels. Papers from the 37th Meeting of the Chicago Linguistic Society.* Vol. 2. Chicago: Chicago Linguistic Society.

Mielke, J. forthcoming. Testing the P-map hypothesis. Ohio State ms.

Nagy, N., & W. Reynolds. 1997. Optimality Theory and variable word-final deletion in Faetar. Language Variation and Change 9:37-55.

Newmeyer, F. 2001. Where is Functional Explanation? in Andronis, Mary, Christopher Ball, Heidi Elston and Sylvain Neuvel eds. *CLS 37: The Panels. Papers from the 37th Meeting of the Chicago Linguistic Society*. Vol. 2. Chicago: Chicago Linguistic Society.

Ohala, J. 1992. Alternatives to the Sonority Hierarchy for Explaining Segmental Sequential Constraints. *Chicago Linguistics Society: Papers from the Parasession on the Syllable*. Chicago: CLS. 319-338.

Ovcharova, O. 1999. A perception-based study of consonant deletion in Turkish. Poster presented at the ICPhS satellite meeting “The Role of Speech Perception Phenomena in Phonology”.

Sezer, E. 1986. An autosegmental analysis of compensatory lengthening in Turkish. in L. Wetzels & E. Sezer, eds. *Studies in Compensatory Lengthening*. Dordrecht. Foris Publications.

Steriade, D. 1997. Phonetics in Phonology: The Case of Laryngeal Neutralization. ms. UCLA.

Steriade, D. 2001a. Directional asymmetries in place assimilation: a perceptual account. In E. Hume and K. Johnson (eds.) The Role of Speech Perception in Phonology. New York: Academic Press.

Steriade, D. 2001b. The Phonology of Perceptibility Effects: the P-map and its consequences for constraint organization. UCLA ms.

Valdman, A. 1976. *Introduction to French Phonology and Morphology*. Newbury House Publishers. Rowley, Massachusetts.

Volaitis, L. E., and J. L. Miller. 1992. Phonetic prototypes: Influence of place of articulation and speaking rate on the internal structure of voicing categories. *JASA* 92 (2), pt. 1.

Winer, B. J. 1971. Statistical Principles in Experimental Design. McGraw-Hill: New York.

Wright, R. 1996. Consonant Clusters and Cue Preservation in Tsou. PhD dissertation. UCLA.

Asymmetries in Speech Errors and their Implications for Underspecification

Marianne Pouplier and Louis Goldstein
Haskins Laboratories and Yale University

1 Introduction

In speech production research, speech errors have long been appealed to in evaluating the psychological reality of linguistic units. The systematic nature of errors with respect to their occurrence and distribution (cf. for instance Fromkin, 1971; Shattuck-Hufnagel & Klatt, 1979; Dell, 1986; Stemberger, 1991a) allows us to assume that those units that appear to behave independently in speech errors are presumably - at least at some point - units of processing. More specifically, it has been proposed that the asymmetric distribution of single segment errors involving coronals provides evidence for their underspecification (Stemberger, 1991a; Stemberger & Stoel-Gammon, 1991).

However, an important caveat in speech error research is that the method for detecting errors relies on impressionistic transcription. This leaves open the possibility that the inherently segmental nature of transcription carries its own bias into the data. Subphonemic errors or errors resulting in a phonologically ill-formed utterance, for instance, are difficult to transcribe in a segmental system, which may be one of the reasons why this kind of error is so rarely reported (cf. among others Fromkin, 1971; Cutler, 1981; Ferber, 1991; Boucher, 1994). In recent years, studies have begun to investigate speech errors by means of instrumental measurements. Articulatory and acoustic studies (Mowrey & MacKay, 1990; Boucher, 1994; Goldstein et al., in prep.; Frisch & Wright, in press) have shown empirically that errors are not a matter of all or nothing - that is, systematic errors can occur below the level of a segment instead of being confined to a temporal misselection of phonological (abstract) units.

This paper follows a new perspective on speech errors within the framework of Articulatory Phonology, as proposed by Goldstein et al. (in prep.). On the basis of kinematic evidence, their work has demonstrated that speech errors are not restricted to categorical exchanges of position of segmental units, but rather gestures that compose segments can exhibit errors that vary from zero to maximal in magnitude.

Here we report results from two perceptual experiments which use stimuli selected on the basis of their articulatory properties only, covering a range of errorful gestural activations. The outcome of the perceptual experiments suggests that different segments show different degrees of vulnerability to (subsegmental) speech errors: While listeners detected errors reliably for some segments, for other segments the reaction to errorful and non-errorful tokens was not distinct. The data suggest that at least for some error types an asymmetric error distribution arises due to perception, while production itself is not asymmetric. However, for error types involving segments whose gestural compositions stand in a subset relationship to each other (as described below), asymmetries may indeed originate in production due to the overall dominance of a gestural intrusion bias observed in the production data of Goldstein et al. (in prep.).

2 Errors as abstract segment exchanges

The examination of corpora of spontaneous errors as well as results of elicitation experiments have shown that the most commonly occurring errors involve single segments, while single feature errors are rarely reported (cf. for instance Fromkin, 1971; Shattuck-Hufnagel, 1979; 1983; Dell, 1986). Likewise syllables are usually not affected by exchange errors. The notion that phonological segments are primary units of word-form retrieval has thus gained wide-spread acceptance, as has the view that errors are phonologically well formed: An activated (abstract) segment is categorically shifted to a wrong position within a 'prosodic frame'. In this new position, the segment will be produced 'normally', as if it were the intended segment. Thus allophonic features typically pertain to the new ('wrong') position of a segment (Shattuck-Hufnagel & Klatt, 1979; Shattuck-Hufnagel, 1983).

Errors show further a frequency bias in that more frequent elements are less likely to be affected by error than less frequent elements. Frequency of occurrence is also reflected in a directionality effect: less frequent elements are usually replaced by more frequent elements. For some segments, however, an anti-frequency bias has been reported (in experimental studies as well as in corpora). While /t/ is a more frequent segment in English than /k/, for instance, and thus the expected substitution-directionality is */k/ $\rightarrow$ /t/, the opposite is actually observed: /t/ is more often substituted by /k/ than vice versa. This is also the case for /s/ and /ʃ/, with /s/ as the more frequent element turning more often into /ʃ/ than vice versa. At the same time, Stemberger (1991a) and Stemberger and Treiman (1986) identify an addition bias in cluster environments: In errors, it is more usual for a segment to be added than to be deleted. Stemberger (1991a) reinterprets the anti-frequency effect as surface manifestation of the addition bias by invoking the concept of coronal underspecification: Given that coronals are underspecified for place of articulation, so he argues, the addition bias will lead other segments' place specifications to intrude more easily (since the 'empty space' is a willing host), independent of segment frequency.

3 Gestures as basic units

The framework of Articulatory Phonology (Browman & Goldstein, 1989; 1992; 2001) opens a new perspective on speech errors and the speech production system. Within Browman and Goldstein's framework, dynamically specified gestures are hypothesized to be the basic units of speech production. Errors can thus be interpreted as reflecting the gestural structure of speech, in that they can activate components of gestural structures in varying magnitudes. That is, both individual gestures as well as larger units consisting of tightly cohesive multiple gestures can be involved in erroneous productions. In a magnetometer (EMMA) study by Goldstein et al. (in prep.) support for these assumptions was gained from kinematic speech error data. Errors were observed to be gradient in nature; that is, an individual gesture can take on a continuum of values, varying from zero to maximal. While a segmental approach would be able to accommodate gradient activations at the articulatory output level, there is evidence for gestures as units of higher levels of organization. It was shown that tightly cohesive multigestural constellations can be broken up in an error, and an individual gesture that forms part of a multigestural unit can isolatedly show up at an erroneous temporal location.[1] Further, asymmetries that occur in the articulatory data cannot stem from abstract

[1] Goldstein et al. (in prep.) define two types of errors, 'reduction' and 'intrusion'. A reduction error is defined as an erroneous decrease of the magnitude of the target constriction (e.g., a decreased magnitude of the tongue tip gesture during /t/). An intrusion error is defined as the constriction of a vocal organ that is not controlled in the normal, non-errorful gestural constellation (e.g., addition of a tongue dorsum gesture during /t/).

segment exchanges: The overall observed bias for gestural intrusion as opposed to reduction has the consequence that often two gestures (one appropriate, one intruding) are produced at the same time. For example, during the repetition of the phrase cop top, errors are observed in which an intruding /k/-like dorsum (TD) gesture is produced concurrently with the tongue tip (TT) gesture of *top*. Goldstein et al.'s (in prep.) results show that there is no asymmetry in production between /t/ and /k/ - the relevant asymmetry is between reduction and intrusion errors instead. This asymmetry affects /t/ and /k/ equally. For /s/ and /ʃ/, the situation is more complex, since the distinction intrusion - reduction can only be defined in terms of the non-shared vocal tract variable (tongue body; TB).[2] For TB, the addition bias is confirmed in Goldstein et al.'s data.

These findings lead to the question where the asymmetry that has been reported between coronals and non-coronals might stem from. From the perspective offered in this paper, asymmetries have two potential sources: Asymmetries may originate in *production*: The intrusion bias may lead to one segment prevailing over another. However, production asymmetries between /t/ and /k/ are not evidenced in the production data of Goldstein et al. Asymmetries may originate in *perception*: The intrusion bias might have different perceptual consequences for different segments. In recording errors, perceptually more salient errors may thus come to be overrepresented. That is, gradient errors and their interaction with perceptual biases might account for the asymmetries. In order to put this possibility to test, Goldstein et al.'s kinematic data were used as stimuli in a perceptual experiment with the aim of determining how the articulatory error distribution maps onto perception.

4 Experiment 1: /t/-/k/

4.1 Method

Goldstein et al. (in prep.) used bisyllabic alternating phrases in their articulatory study (e.g. "coptop"; "kiptip"), which were repeated by the subject for about 10 seconds at a time. They distinguish between errorful and non-errorful utterances as follows: Their control condition involved non-alternating phrases (e.g. "copcop", "toptop"), during which no errors occurred. Error definition will be exemplified here for tongue dorsum (TD) magnitude during a /t/ from an alternating trial. The tongue dorsum value during the intended /t/ from the alternating condition is evaluated against the distribution of the non-alternating controls. If the token's gestural magnitude is more than two standard deviations from the TD mean of the /t/-control, it is classified as an 'error'. Further, gradient and categorical errors are distinguished: A gradient error on TD during /t/ is more than two standard deviations from the /t/-control mean, but not as extreme (less than two standard deviations) from the TD values of the /k/ control mean. A categorical error on this particular /t/ token would involve a gestural magnitude which is not only more than two standard deviations from the /t/ control mean but even exhibits a TD value that is within the two standard deviation range of the TD values of the /k/ controls. That is, the TD value of a /t/ defined as categorical error is within the 96% range of the distribution of a non-errorful /k/.

The digitized audio of theoriginal bisyllabic utterances of the production experiment were edited into single syllable utterances (e.g., "cop"). In a simple (not choice) reaction time task, participants were instructed to listen to "short words" presented in random order and

[2] Browman and Goldstein (2001) hypothesize the gestural specification for /ʃ/ to involve TT and TB, while the gestural specification of /s/ involves TT only. Since an intrusion error is defined by Goldstein et al. (in prep.) as intrusion of a constriction that is not controlled in the non-errorful gestural constellation, this criterion does *per definitionem* not apply to TT during either /s/ or /ʃ/.

decide whether the words begin with a particular consonant sound. Subjects ranged from 18 to 44 years of age and were paid for their participation. 14 subjects were tested. Data from 3 subjects were discarded since their identification rate for the error-free controls was below 50%.

The stimulus list contained 71 single syllable tokens which were selected according to their membership in a particular 'error category' (cf. Table 1).[3]

	categorical	**gradient**	**no error**
Intrusion	*t* (10)	*t* (10)	
	k (12)	*k* (8)	*t* (10)
Reduction	-	*t* (5)	*k* (10)
	-	*k* (6)	

Table 1. Single syllable stimuli for perceptual task grouped into error categories. Numbers in brackets indicate the number of tokens representing each category.

None of the tokens selected has an error on more than one constriction (i.e. never a reduction as well as intrusion error on the same token). Where possible, a minimum of 10 tokens from each category was selected for both /t/ and /k/. Two major categories of exceptions were categorical and gradient reduction errors: The data collected in Goldstein et al.'s (in prep.) EMMA experiment did not contain any reduction errors of categorical magnitude in TD during /k/ with no accompanying error on TT. Likewise there are no categorical reduction errors on TT during a /t/ with no error simultaneously occurring in TD. Also gradient reduction errors, i.e. errors on the target gesture (TD for /k/ and TT for /t/) with no errorful intruding gesture (TD for /t/ and TT for /k/) are also underrepresented with only one occurrence for *kip* and *tip*, but 4 during *top* and five during *cop*.[4]

Using Psyscope, the stimuli were presented 12 times overall (6 times each in the 2 different conditions described below), randomized differently each time. Subjects sat in a soundbooth in front of a computer screen and a button box. Stimuli were presented over headphones. Two different monitoring conditions were employed: Subjects were asked to decide whether they heard an initial /t/-sound (condition 1), or an initial /k/-sound (condition 2). If they heard the given sound, they were instructed to press a response button as quickly as possible, otherwise, they were instructed to wait for the next trial. The conditions were blocked in cycles of 3, that is, the program cycled 3 times through the entire /t-k/ stimulus list in 3 different randomizations while subjects monitored for /t/. In the subsequent 3 cycles, subjects were asked to monitor for /k/, then again for /t/ and once more for /k/. Between these blocks of 3 cycles, subjects were given the option to take self-terminated breaks, i.e. subjects end breaks by pressing the response button. During a given block, a letter representing the sound subjects should be monitoring for was displayed on the screen. The time between the onset of two successive stimuli was 2000 ms, partitioned into a 1500 ms response window and 500 ms inter-trial time. During the window of 1500 ms subjects heard the audio stimulus and response time was measured (the response window started with the onset of the audio stimulus). Independent of whether a response was recorded, the next trial came up after 2000 ms.

[3] Goldstein et al.'s study included various rate, stress, vowel (*cop top* and *kip tip*), and phrase position conditions. The tokens selected for the present experiment were distributed across all of these conditions.

[4] The difference between /t/ and /k/ in the gradient-categorical intrusion categories is due to a coding mistake in the experimental setup; these numbers do not reflect genuine differences in frequency of error.

4.2 Results and discussion

The goal of our analysis was to determine how detectability of /t/ and /k/ was affected by error status. Instead of analyzing percent of correct identifications directly, the data were transformed using a non-parametric test for sensitivity (Grier, 1971). This sensitivity measure takes into account subjects' inherent response bias by adjusting the number of 'hits' (i.e. correct identification responses) for the number of 'false alarms' (i.e. incorrect positive responses). The adjustment formula, given in (2), yields a maximal identification value of 1 (only hits, no false alarms) and a minimum value of 0 (only false alarms):

(2) I = (1-P(fa) + P(hit))/2

where P(hit) = P(r|G) and P (fa) = (r|X) and r is a response. G is a stimulus with target gesture r, and X is a stimulus not having target gesture r.

Results are shown in Table 4. A 2-way ANOVA with repeated measure on both factors was performed. The two factors are error type (categorical - gradient - no error) and intended target (t-k), whereby intended target refers to the speaker's target in Goldstein et al.'s production study. Each subject contributes one sensitivity value per intended target per error type. Factor error type is significant at $p < 0.01$ ($F(2, 20) = 51.13$, $p < 0.0001$); factor intended target is not significant ($F(1, 10) < 1$), indicating that the overall sensitivity is the same for /t/ and for /k/. The interaction effect is significant ($F(2, 20) = 21.83$; $p < 0.0001$). The interaction arises because the effect of error category is stronger for /t/ than for /k/.

error type	intended target *t*	intended target *k*
categorical	0.73 (0.11)	0.85 (0.07)
gradient	0.9 (0.04)	0.86 (0.07)
no error	0.97 (0.03)	0.86 (0.08)

Table 2. Means (and standard deviations) for sensitivity results for /t/ and /k/ grouped by error category.

A 1-way ANOVA follow-up with a posthoc (Ryan-Einot-Gabriel-Welsch Multiple Range Test) on the interaction means shows that while for /t/ all three error types are significantly different from each other, none of the means for /k/ differ significantly. These results indicate that errors are asymmetric in that /t/ is perceptually more affected by error than /k/.

As to the apparent anti-frequency bias that has been reported for /t/ and /k/ for error data collected by means of transcription, these results indicate that these asymmetries reflect a property of the perceptual system rather than the production system: if errors are systematically heard more easily on /t/ than on /k/, this perceptual asymmetry will substantially affect the error distribution in corpora.

5 Experiment 2: /s/-/ʃ/

5.1 Subjects and experimental setup

The same experimental setup and the same subjects (in a separately scheduled session) were used for /s/-/ʃ/; 13 of the earlier subjects were available. Due to coarticulation, the release of the preceding /p/-closure in a *sopshop* phrase is audible during the frication, even after the utterances have been cut up into individual syllables. To ensure that subjects would parse the bilabial release as coda instead of as complex /ps/ or /pʃ/ onset, a syllable /op/ was spliced at the beginning of all tokens.[5] A silence interval of 100 ms was spliced to the end of the /o/ vowel to make the /p/ closure of constant length. The instructions specified that subjects would hear bisyllabic words with the first syllable always being /op/. It was specifically pointed out to them that there was no consonant at the beginning of the word. Their task was specified as determining whether the second syllable begins with a given consonant sound.

For Experiment 2, data from 2 subjects were discarded since their identification rate for the error-free controls was below 50%. The stimuli distribution is given in Table 9. Selected tokens are distributed across all rate, stress and phrase position conditions.

TB	**categorical**	**gradient**	**no error**
intrusive	*s* (5)	*s* (4)	*s* (5)
reductive	ʃ (5)	ʃ (5)	ʃ (5)

Table 3. Stimulus categories for /s/-/ʃ/. Numbers in brackets indicate the number of tokens representing each category.

Since only the activity of one tract variable, i.e. tongue body (TB), can be differentiated in terms of reduction and intrusion as defined by Goldstein et al. (cf. fn 2), fewer tokens were tested than in Experiment 1. In addition, only error data for one vowel condition were available from the EMMA experiment. The stimulus list for the fricatives thus contained 29 tokens.

Recall that for /t/-/k/, stimuli were selected such that there are never two co-occurring errors during one token, i.e. no token has an error on both constrictions at the same time. For /s/-/ʃ/ this selection criterion had to be modified, since TT and TB receivers are not independent in the way TT and TD are (TT and TB receivers were about 20 mm apart). That is, it is not possible to select tokens with an error on TT only or TB only; for the vast majority of tokens, an error on one tract variable is accompanied with an error on the other tract variable. Neither is it possible to systematically vary error degree (i.e. gradient TT with categorical TB error, categorical TT with categorical TB error, etc.), since not enough representatives of each type are in the EMMA data. Note that non-errorful tokens are truly 'error free'; gestural magnitudes for both constrictions are well within 1 standard deviation of the control mean.

[5] The syllable was a stressed, fast rate, non-errorful utterance of *shop*: The frication part as well as the first temporal half of the vowel were cut off (resulting in a vowel duration of 55.1 ms).

5.2 Results

Like for /t/-/k/, a 2-way repeated measures ANOVA was performed on the nonparametric sensitivity measure with repeated measures on both factors. Factors are error type (categorical-gradient-no error) and intended target (/s/-/ʃ/). Factor error type is significant ($F(2,22) = 289.04$; $p < 0.0001$). Factor intended target is not significant ($F(1, 11) = 1.74$; $p = 0.2142$); the interaction effect reaches significance ($F(2,22)=15.12$, $p < 0.001$).

error type	s	ʃ
categorical	0.41 (0.14)	0.36 (0.23)
gradient	0.88 (0.14)	0.6 (0.18)
no error	0.92 (0.1)	0.96 (0.04)

Table 4. Means (and standard deviations) for sensitivity results for /s/ and /ʃ/ grouped by error category.

A 1-way ANOVA on the interaction means with a posthoc test (Ryan-Einot-Gabriel-Welsch Multiple Range Test) shows that the sensitivity values for /ʃ/ are significantly different for categorical, gradient and no error. For /s/, on the other hand, only categorical errors significantly affect sensitivity. There is no statistically significant difference between the no error and gradient error categories. These results were unanticipated: the perception of /ʃ/ was found to be more affected by error than was the perception of /s/.

The results for /s/-/ʃ/ show a slight directionality in that the identification of /ʃ/ is more variable under error as the identification of /s/. However, the asymmetry is overall relatively weak, compared to the asymmetry obtained for /t/ and /k/ in Experiment 1. It has to be considered whether the error status of tongue tip affects the outcome of the perceptual results for /s/ and /ʃ/. Defining an error in terms of intrusion and reduction as done in this paper precludes an analysis of TT for /s/ and /ʃ/ analogous to TB. However, errors on TT can be identified for both, /s/ and /ʃ/, since they differ significantly in TT height (/ʃ/ has a higher TT position than /s/). That /s/ and /ʃ/ are equally affected by categorical TB errors might be due to the fact that for these tokens, also the TT gesture is of errorful magnitude. As has been mentioned before, TT and TB do not behave independently in errors due to the close vicinity of the constriction locations. In contrast to /s/, a gestural specification of /ʃ/ further includes an upper lip (UL) protrusion gesture. For the subject whose kinematic data were used for the perceptual experiment, Goldstein et al. (in prep.) could not reliably measure a difference in UL protrusion for /s/ and /ʃ/ in a way that would allow them to statistically determine errorful UL behavior. Nevertheless it cannot be excluded that the presence/absence and magnitude of a UL gesture will interact with perception. The perceptual results for /s/ and /ʃ/ therefore cannot be interpreted in the same way as they can for /t/ and /k/. For the stops, the experiment shows the perceptual consequences of a co-production of two gestures. For the sibilants, the experimental stimuli are less tightly controlled; the experiment shows the effect of the occurrence of at least one errorful constriction. The results nonetheless allow to come to generalizations about asymmetries, since the lack of independence between TT and TB in Goldstein et al.'s (in prep.) experiment suggests that the occurrence of single-constriction errors between these two vocal organs is extremely rare, if not impossible (due to the close vicinity of the constriction locations).

6 General discussion

In the data presented here, asymmetries have been found in production as well as in perception. However, these asymmetries are different in nature from the ones that have been reported by Stemberger (1991a, 1991b). Goldstein et al. (in prep.) show that both /t/ and /k/ exhibit an intrusion bias; there is no production asymmetry between the two stops. An asymmetry does exist in the form of a gestural addition bias, but this phenomenon affects /t/ and /k/ to the same degree. The present experiments demonstrate that it is in perception that the gestural intrusion bias has different consequences. The perception of coronals is more affected by errors than the perception of velars. The anti-frequency effect in substitution errors that has been observed for /t/ and /k/ for error data recorded by means of transcription might thus be due to the transcriber's perceptual biases. Crucially, our data do not require an appeal to coronal underspecification in order to explain the apparent anti-frequency effects in error distributions.

For /s-ʃ/ on the other hand, the perceptual asymmetry cannot explain the directionality effect found in speech error corpora and experiments. On the basis of the perceptual results alone a directionality effect to be expected would be /ʃ/ → /s/ not */s/ → /ʃ/. Since /ʃ/ is more systematically affected by error than /s/, we would expect /ʃ/ to be more often substituted by /s/. This, however, is the opposite of the asymmetry that has been recorded in speech error research. This suggests that the asymmetries cannot be explained by perceptual biases, they must originate in production. Recall that in production, intrusive errors are more frequent than reductive errors. For the sake of clarity, the hypothesized gestural composition of /s/ and /ʃ/ shall be repeated here schematically:

/ʃ/: TT and TB gesture
/s/: TT gesture only

Errors in production are dominated by the intrusion bias. For /ʃ/, there is no intrusion bias in terms of TT, since both TT and TB are actively controlled for in normal production. Thus the most common error should be an intrusive TB error on /s/. This means that /s/ will systematically be more affected by errors compared to /ʃ/. The data obtained in the production experiment confirm this prediction.

It can be concluded that asymmetries can originate in production where the interacting segments are in a subset relationship to each other. Thus intrusive TB errors only affect /s/, not /ʃ/. Note that this supports a more limited notion of underspecification assumed in a gestural framework: gestures have task specific targets for certain tract variables only, and segments typically correspond to sets of gestures that leave several tract variables unspecified (Browman & Goldstein, 1992). Again, there is no need to assume /s/ to be underspecified for place of articulation; it does lack, however, a tongue body constriction gesture. It is to be expected that the 'palatal bias' reported in Stemberger (1991a), in which /t/ turns more often into the affricate /tʃ/ than vice versa, can be explained on the same basis.

The subset relationship between /s/ and /ʃ/ in terms of their gestural composition is an independently motivated assumption within Articulatory Phonology (Browman and Goldstein, 2001) and does not hold of their standard featural differentiation (+/- anterior). The data presented here can be taken as supporting evidence for their asymmetric gestural composition.

7 Conclusions

The experiments presented in this paper provide evidence for systematic asymmetries in the perception of speech production errors. These asymmetries combine with the nature of production errors to explain the patterns of asymmetries reported in speech error corpora. In production, error patterns are generally dominated by an intrusion bias: it is more likely for an errorful gesture to intrude than it is for a target gesture to be reduced. For different segments, this gestural addition bias produces different consequences. For /t/-/k/, intruding TD gestures during /t/ have a systematic perceptual effect, whereas for /k/ intruding TT gestures do not significantly affect identification or reaction times. For /s-ʃ/, perceptual biases are not the source of distributional asymmetries. Rather, the addition bias translates into a 'phoneme bias'. The most likely error to occur is an intruding TB gesture during /s/; the intrusion bias leaves /ʃ/ unaffected by /s/ since /ʃ/ and /s/ are gesturally in a subset relationship to each other. The study shows that the concept of coronal underspecification is not needed to explain asymmetries in speech errors. The data further support the notion that gestures are units of speech production, since the obtained results can be accounted for by the specific assumptions Articulatory Phonology makes about the gestural composition of segments.

Acknowledgments

Thank you to Larissa Chen for comments and much help in practical matters. Thank you also to Alexei Kochetov, Dani Byrd, Donca Steriade, Ken de Jong, and Stefanie Shattuck-Hufnagel for insightful comments and discussions. Work supported by NIH grant HD-01994.

References

Boucher, Victor J. (1994): Alphabet-related biases in psycholinguistic enquiries: considerations for direct theories of speech production and perception. *Journal of Phonetics, 22*(1), 1-18.

Browman, Catherine P., & Goldstein, Louis (1989): Articulatory Gestures as Phonological Units. *Phonology, 6*(2), 201-251.

Browman, Catherine P., & Goldstein, Louis (1992): Articulatory Phonology: An Overview. *Phonetica: International Journal of Speech Science, 49*, 155-80.

Browman, Catherine P., & Goldstein, Louis (2001): *Articulatory Phonology*. Manuscript.

Cutler, Anne (1981): The reliability of speech error data. *Linguistics, 19*, 561-582.

Dell, Gary (1986): A Spreading-Activation Theory of Retrieval in Sentence Production. *Psychological Review, 93*(3), 283-321.

Ferber, Rose (1991): Slip of the Tongue or Slip of the Ear? On the Perception and Transcription of Naturalistic Slips of the Tongue. *Journal of Psycholinguistic Research, 20*(2), 105-122.

Frisch, Stefan, & Wright, Richard (in press). The phonetics of phonological speech errors: An acoustic analysis of slips of the tongue.

Fromkin, Victoria A. (1971): The Non-Anomalous Nature of Anomalous Utterances. *Language, 47*, 27-52.

Goldstein, Louis, Pouplier, Marianne, Chen, Larissa, & Byrd, Dani (in preparation). Gestural Action Units Slip in Speech Production Errors.

Grier, J. Brown (1971): Nonparametric indexes for sensitivity and bias: Computing formulas. *Psychological Bulletin, 75*, 424-429.

Mowrey, Richard A., & MacKay, Ian R. (1990): Phonological Primitives: Electromyographic Speech Error Evidence. *Journal of the Acoustical Society of America*, *88*(3), 1299-1312.

Shattuck-Hufnagel, Stefanie (1979): Speech Errors as Evidence for a Serial-Ordering Mechanism in Sentence Production. In W.E. Cooper & E.C.T. Walker (Eds.), *Sentence Processing: Psycholinguistic Studies Presented to Merrill Garrett*. Hillsdale, NJ: Lawrence Erlbaum. 295-342.

Shattuck-Hufnagel, Stefanie (1983): Sublexical Units and Suprasegmental Structure in Speech Production Planning. In P.F. MacNeilage (Ed.), *The Production of Speech.* New York: Springer. 109-136.

Shattuck-Hufnagel, Stefanie, & Klatt, Dennis (1979): The limited use of distinctive features and markedness in speech production: Evidence from speech error data. *Journal of Verbal Learning and Verbal Behavior, 18*, 41-55.

Stemberger, Joseph, & Stoel-Gammon, Carol (1991): The Underspecification of Coronals: Evidence from Language Acquisition and Performance Errors. In C. Paradis & J.-F. Prunet (Eds.), *The Special Status of Coronals. Internal and External Evidence.* San Diego: Academic Press. 181-199.

Stemberger, Joseph, & Treiman, Rebecca (1986): The Internal Structure of Word-Initial Consonant Clusters. *Journal of Memory and Language, 25*, 163-180.

Stemberger, Joseph Paul (1991a): Apparent Anti-frequency Effects in Language Production: The Addition Bias and Phonological Underspecification. *Journal of Memory and Language, 30*, 161-185.

Stemberger, Joseph Paul (1991b): Radical underspecification in language production. *Phonology, 8*, 73-112.

Patterns of Jaw Coarticulatory Direction and Linguomandibular Coordination in VCV Sequences

Daniel Recasens
Institut d'Estudis Catalans and
Universitat Autònoma de Barcelona

1 Introduction

Data on lingual movement, dorsopalatal contact and F2 frequency presented in previous papers of ours (Recasens, 2002; Recasens and Pallarès, 2001; Recasens, Pallarès and Fontdevila, 1997) suggest that the degree of articulatory constraint (DAC) model accounts to a large extent for the extent and direction of tongue dorsum coarticulation in VCV and CC sequences. A goal of this investigation is to verify the predictions of this model with respect to jaw V-to-V effects in VCV sequences using articulatory movement data collected with electromagnetic articulometry (EMA).

1.1 The DAC model

According to the DAC model of coarticulation, consonants may be more or less constrained based on demands on place and manner of articulation. Some consonants appear to be specified for a high degree of articulatory constraint, i.e., dorsals (in agreement with the sluggishness of the tongue dorsum and the large contact area involved in their production), lingual fricatives and trills (in line with the precise mechanisms used by speakers for their implementation, namely, tongue grooving for the former and adequate tongue shape and elasticity for the latter), and dark /l/ (since this realization requires the formation of a secondary postdorsal constriction at the upper pharynx). Regarding consonants specified for a lower degree of tongue dorsum constraint, coupling between the tongue dorsum and the primary tongue front articulator causes dentals and alveolars to be more constrained than bilabials. Differences in tongue predorsum activation render /i/ more constrained than /a/ and /u/, and /ə/ more unconstrained than the other vowels.

The degree of constraint for the intervocalic consonant has important consequences for the extent and direction of the coarticulatory effects in VCV sequences. Indeed, the DAC value for the consonant is inversely related to the degree of V-to-C coarticulation (i.e., to how sensitive the consonant is to the influence of the adjacent vowels) and positively related to the C-to-V effects (i.e., to how much the consonant affects the adjacent vowels). Moreover, as shown below, requirements on consonantal production may also explain whether C-to-V anticipation prevails over C-to-V carryover, or viceversa. Trends in V-to-V direction may also be accounted for assuming that vocalic anticipation ought to vary inversely with the prominence of the consonantal carryover effects while vocalic carryover effects are expected to decrease with the strength of the consonant-dependent anticipatory component. The following patterns of C-to-V and V-to-V coarticulatory direction may be identified:

(a) Among consonants with a high DAC value, dark /l/ favors C-to-V1 anticipation over C-to-V2 carryover consistently with the fact that tongue dorsum lowering and retraction often starts before tongue tip raising for the implementation of this consonantal realization. On the other hand, consonants involving tongue dorsum raising such as alveolopalatals and velars favor C-to-V2 carryover effects in accordance with the slow motion of the primary dorsal articulator at consonantal release which may be due to inertia.

In agreement with these C-to-V patterns, data reported in our previous papers reveal that VCV sequences with dark /l/ allow more vocalic anticipation than vocalic carryover while those with the alveolopalatal /ɲ/ show more vocalic carryover than vocalic anticipation. A specific situation applies to VCV sequences with dorsovelars: vocalic effects from /i/ vs /a/ favor the expected carryover direction when the transconsonantal vowel is /i/ (e.g., in the sequence pair /aki/-/iki/) but the anticipatory direction when the transconsonantal vowel is /a/ presumably since forward tongue dorsum motion during the velar closure period causes extensive vocalic anticipation to occur (e.g., in the sequence pair /aki/-/aka/).

(b) Two basic patterns of C-to-V direction are found in VCV sequences with dentals and alveolars but for dark /l/ (and, to some extent, for labials as well). Consonantal effects on /a/ happen to be more prominent at the anticipatory level presumably since this vowel permits free apical anticipation; on the other hand, C-to-V effects on /i/ are especially salient at the carryover level since this vowel contributes to tongue dorsum raising during the consonant. Accordingly, vowel effects appear to favor the carryover component in the latter scenario (e.g., in the sequence pair /ati/-/iti/) and the anticipatory direction in the former one (e.g., in the pair /ati/-/ata/).

1.2 Jaw coarticulation

Data on jaw coarticulation are scarce. To a large extent, trends in C-to-V coarticulation for the jaw resemble those for the tongue dorsum in the light of evidence showing that the former articulatory structure may assist the latter along the height dimension (e.g., high vowels involve both tongue body and jaw raising while low vowels are produced with a lowered tongue body and jaw). Indeed, previous studies addressing this issue (Keating, Lindblom, Lubker and Kreiman, 1994) reveal that high vowels are more resistant than low vowels to C-to-V effects in jaw height, and that such effects become more prominent as a function of consonants involving a higher jaw (dentoalveolars including /s/ but not so /l/) than of those produced with a lower jaw (labials, /l/, velars). Compatibly, consonants of the latter group are more likely to be influenced by V-to-C effects in jaw height than those of the former group.

Regarding coarticulatory direction, data in the literature reveal that the jaw often favors the carryover over the anticipatory component both for the consonantal and for the vocalic effects which could be associated with the relatively slow motion of the mandible in speech. This trend is documented in VCV sequences with labial consonants, both in the light of articulatory data (Tuller, Harris and Gross, 1981; Sussman, MacNeilage and Hanson, 1973) and of acoustic data on F1 coarticulatory effects from /i/ vs /a/ in English /VbəbV/ sequences (Magen, 1997). (F1 is known to reflect changes in mouth opening degree and should thus be correlated with variations in jaw height). In a recent F1 coarticulation study with the same consonants included in the present investigation (Recasens and Pallarès, 2000), C-to-V and V-to-V effects from /i/ vs /a/ in VCV sequences with transconsonantal /a/ (e.g., /iCa/-/aCa/) were also found to be more prominent at the carryover level than at the anticipatory level; on the other hand, V-to-V effects in sequences with transconsonantal /i/ (e.g., /iCa/-/iCi/) were reported to favor the anticipatory direction in sequences with lingual fricatives, and the carryover direction in those with consonants produced with a low jaw (dark /l/, velars). Other studies report however considerable anticipatory and carryover vocalic effects in jaw coarticulation in VCV sequences with bilabial, alveolar and velar stops (Fletcher and Harrington, 1999).

2 Method

2.1 Recording procedure

Three Catalan speakers uttered ten times all possible VCV combinations with the consonants /p, n, l, s, ʃ, ɲ, k/ (where /l/ is dark and /ɲ/ is an alveolopalatal) and the vowels /i, a, u/ embedded in the Catalan sentence ['gɾaβə pV'CVp ə'βans] ("He records pVCVp earlier"). The inclusion of contextual labial consonants and a schwa ensured that the vowel-dependent coarticulatory effects of interest could be expanded sufficiently along the time domain.

Articulatory movement and acoustic data were collected simultaneously using electromagnetic articulometry by means of a Carstens Articulograph system AG-100. This system consists of a head mount with three magnetic transmitters that generate a magnetic field, and a set of small transducer coils that can be attached to different articulatory structures in the midsagittal plane. As the articulators move inside the vocal tract, the transducer coils induce a signal that is inversely proportional to the cube of the distance between transmitter and transducer. The resulting signal results in a set of voltages which can be converted to distance. In the present experiment coils were placed on the tongue tip, tongue blade, tongue dorsum, incisors of the lower jaw, and upper and lower lip, as well as on the bridge of the nose and upper incisors for head movement correction. Estimates for the subjects' occlusal planes were obtained as anatomical references to which the data could be rotated, as well as traces of their palates.

Movement and acoustic data were digitized using a real-time input system at a sampling rate of 250 Hz for movement and 10 kHz for speech; the time resolution of the EMA data was 4 ms. The kinematic data were converted from voltage to distance, corrected for head movement, rotated to the occlusal plane, and extracted into separate articulatory channels for the X and Y dimensions.

2.2 Data analysis

For each VCV repetition, the onset and offset of the intervocalic consonant were identified from spectrographic and waveform displays, and occasionally from the movement data according to criteria summarized elsewhere (Recasens, 2002). The temporal extent of V-to-V coarticulatory effects was analyzed from the onset of $[\beta]_1$ to the offset of $[\beta]_2$ in the sequence [βə##pVCVp##əβ].

Vocalic temporal effects were considered to occur as long as a significant vowel-dependent difference in articulatory displacement (i.e., /i/ vs /a/, /i/ vs /u/ and /a/ vs /u/, referred to as 'changing' vowel in this paper) extends into the consonant and the transconsonantal vowel (i.e., /i/, /a/ and /u/, referred to as 'fixed' vowel from here forwards). Significant differences as a function of changing vowel for each consonant and fixed vowel condition were computed for X and Y movement data for all six articulatory regions TT, TL, TD, J, UL and LL: results for TDX, TDY and TTX were presented in Recasens (2002) and those for JY will be given in the present paper. V-to-V mandibular effects were measured in jaw height only (i.e., for changing /i/ vs /a/ and /a/ vs /u/) considering the close relationship between jaw and tongue dorsum elevation and the fact that jaw horizontal movement is often small in speech production.

In order to carry out the statistical evaluation of interest, mean articulatory trajectories across repetitions for a given asymmetrical sequence (e.g., effects from V2= /i/ on /ɲ/ and V1=/a/) were compared with those in symmetrical sequences composed of the same consonant and the same fixed transconsonantal vowel (i.e., /aɲa/). One-way ANOVAs

Scheffé ($p< 0.05$) were applied every 4 ms starting at consonantal offset back to $[\beta]_1$ onset in order to determine the extent of vocalic anticipation and from consonantal onset until $[\beta]_2$ offset in order to estimate the extent of vocalic carryover. ($[\beta]_1$ onset and $[\beta]_2$ offset were identified with the shortest temporal values for a given pair of symmetrical and asymmetrical VCV sequences across repetitions). The last significant difference counting backwards during fixed V1 was taken to be the onset of a V2-dependent anticipatory effect and the last significant difference counting forwards during fixed V2 was taken to be the offset of a V1-dependent carryover effect.

Significant coarticulation times obtained according to the procedure just described are shown in Table I across speakers. Those values were submitted to further statistical analysis in view of the large speaker-dependent variability involved (see standard deviations in the table). Repeated measures ANOVAs and post-hoc tests (Scheffé) were performed with speaker as a factor and coarticulation time as the dependent variable ($p< 0.05$). Two analyses were carried out. In test 1, main effects and interactions for JY were computed for the independent variables 'direction' (anticipatory, carryover), 'changing vowel' (/i/ vs /a/, /a/ vs /u/) and 'consonant' (/p, n, l, s, ʃ, ɲ, k/). In order to evaluate the role of the fixed vowel in the duration of the coarticulatory effects, test 2 was performed for the variables 'direction', 'consonant' and 'fixed vowel' separately for changing front /i/ vs back /a/ (test 2a) and for changing low back /a/ vs high back /u/ (test 2b). Significant effects for the mean values of interest are presented in Table II.

3 Results

3.1 Coarticulatory durations

(a) JY data yielded no significant main effects but two significant interactions according to results from test 1, i.e., changing vowel x direction ($F(2,2)=4.108$, $p<0.052$) and consonant x direction ($F(2,12)=4.227$, $p< 0.004$). The former interaction is associated with differences in carryover duration for /i/ vs /a/ (189 ms) > /a/ vs /u/ (127 ms), and the latter interaction with longer effects for /ɲ/ than for /s/ also at the carryover level. Coarticulatory durations in the consonant x direction panels of Table II reveal indeed that, in comparison to the other consonants, those for /ɲ/ are longer at the carryover level while those for /s/ are shorter at the carryover level and longer at the anticipatory level.

These coarticulation trends are similar to those for TDY (correlation coefficients between the coarticulatory durations for JY and for TDY yielded an overall r value of 0.57). Indeed, a comparison between the dotted and continuous lines at the right bottom graph of Figure 1 reveals similar V-to-V coarticulatory durations both for TDY and JY in the case of VCV sequences with dorsal /ɲ/ and, less so, in those with /k/ (correlation coefficients were 0.92 for the former sequences and 0.74 for the latter). Moreover, a more detailed inspection of the V-to-V coarticulatory durations for both dorsal consonants reveals that they are longer in sequences with /i/ than in those without /i/. Vocalic effects for the fricatives yielded lower correlation values between TDY and JY (r values were 0.74 for /s/ and 0.52 for /ʃ/) and, as shown by the left bottom graph in the figure, those in sequences without /i/ were often longer than those in sequences with /i/. Correlation values were also high for /p/ (0.80) but not so for the alveolars /n/ and /l/ (see top graph of Figure 1).

Table I. Temporal extent of significant V-to-V anticipatory and carryover effects for JY across speakers. Data (in ms) are listed as a function of consonant, and changing and fixed vowel condition. Standard deviations are also given in italics.

	Changing i vs a		**Changing a vs u**	
	(fixed i)	*(fixed a)*	*(fixed a)*	*(fixed u)*
p *(Ant)*	**168**	**169**	**256**	**64**
	156	*96*	*118*	*46*
p *(Car)*	**139**	**245**	**164**	**35**
	140	*175*	*80*	*33*
n *(Ant)*	**173**	**137**	**151**	**127**
	150	*103*	*73*	*134*
n *(Car)*	**240**	**228**	**132**	**112**
	170	*139*	*136*	*161*
l *(Ant)*	**185**	**137**	**173**	**175**
	145	*130*	*141*	*147*
l *(Car)*	**191**	**131**	**121**	**100**
	149	*169*	*104*	*52*
s *(Ant)*	**220**	**199**	**167**	**368**
	173	*155*	*176*	*11*
s *(Car)*	**115**	**89**	**20**	**117**
	199	*144*	*22*	*151*
ʃ *(Ant)*	**253**	**197**	**263**	**231**
	188	*174*	*191*	*175*
ʃ *(Car)*	**179**	**165**	**217**	**16**
	187	*178*	*168*	*7*
ɲ *(Ant)*	**92**	**240**	**168**	**121**
	114	80	177	129
ɲ *(Car)*	**335**	**287**	**261**	**235**
	31	*127*	*171*	*8*
k *(Ant)*	**251**	**181**	**195**	**213**
	84	*86*	*134*	*127*
k *(Car)*	**131**	**173**	**172**	**75**
	145	*178*	*118*	*91*

Table II. Mean values across speakers (in ms) and significant differences (in brackets) for JY vowel-to-vowel coarticulation times. Data for the consonant x direction condition are plotted as a function of changing vowel (columns) and consonant and direction (rows). Data for the consonant x fixed vowel condition are presented as a function of changing vowel (columns) and consonant and fixed vowel (rows).

Consonant x direction

	(i vs a; a vs u)	**(i vs a)**	**(a vs u)**
p *(Ant)*	164	169	160
p *(Car)*	146	192	99
n *(Ant)*	147	155	139
n *(Car)*	178	234	122
l *(Ant)*	168	161	174
l *(Car)*	136	161	111
s *(Ant)*	238	209	267
s *(Car)*	85	102	69
ʃ *(Ant)*	236	225	247
ʃ *(Car)*	144	172	117
ɲ *(Ant)*	155	166	145
ɲ *(Car)*	279	311	248
k *(Ant)*	210	216	204
k *(Car)*	138	152	123

C x fixed V

	(i vs a)
p *(fixed i)*	153
p *(fixed a)*	207
n *(fixed i)*	207
n *(fixed a)*	183
l *(fixed i)*	188
l *(fixed a)*	134
s *(fixed i)*	167
s *(fixed a)*	144
ʃ *(fixed i)*	216
ʃ *(fixed a)*	181
ɲ *(fixed i)*	213
ɲ *(fixed a)*	263
k *(fixed i)*	191
k *(fixed a)*	177

	(a vs u)
p *(fixed a)*	210
p *(fixed u)*	49
n *(fixed a)*	141
n *(fixed u)*	119
l *(fixed a)*	147
l *(fixed u)*	137
s *(fixed a)*	93
s *(fixed u)*	243
ʃ *(fixed a)*	240
ʃ *(fixed u)*	123
ɲ *(fixed a)*	215
ɲ *(fixed u)*	178
k *(fixed a)*	183
k *(fixed u)*	144

(b) According to the two consonant x fixed panels in Table II, V-to-V coarticulatory effects associated with changing /a/ vs /u/ were longer for fixed /a/ than for fixed /u/ for most consonants (right panel) while those associated with changing /i/ vs /a/ were generally longer when the fixed vowel was /i/ than when it was /a/ (left panel). Differently from these JY effects, V-to-V effects for TDY (see Recasens, 2002) were consistently longer in the fixed /a/ and /u/ condition than in the fixed /i/ condition.

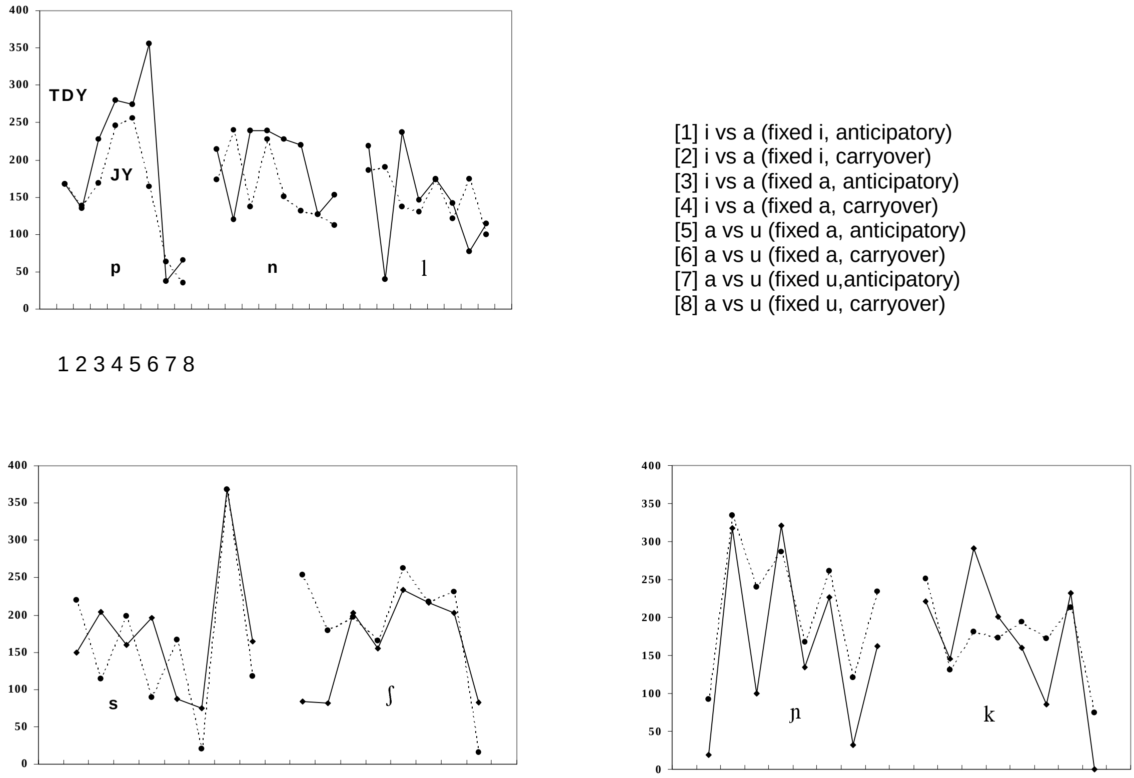


Figure 1. V-to-V coarticulatory durations for TDY and JY. Effects are displayed as a function of the intervocalic consonant in the changing and fixed vowel conditions.

3.2 Coarticulatory direction

Figure 2 displays differences in V-to-V coarticulatory direction for JY across speakers. Data on V-to-V coarticulatory direction for TDY are also given for comparison (see Recasens, 2002). In order to obtain those differences, the vowel-dependent carryover effects were subtracted from the vowel-dependent anticipatory effects for each consonant and all pairs of changing vowels in each fixed vowel context condition. Bars in the figure plot differences between carryover and anticipatory effects for changing /i/ vs /a/ in the fixed /i/ and /a/ conditions (dark bars), and for changing /a/ vs /u/ in the fixed /a/ and /u/ conditions (white bars).

The JY coarticulatory effects appear to favor the same direction as the TDY effects in VCV sequences with dorsal consonants /ɲ/ and /k/, i.e., vocalic carryover for /ɲ/ and vocalic anticipation for /k/. Regarding VCV sequences with the lingual fricatives /s/ and /ʃ/, JY effects are anticipatory rather than carryover, and the relative prominence of the anticipatory vs carryover effects is more obvious for JY than for TDY mostly so in the fixed vowel /i/ condition.

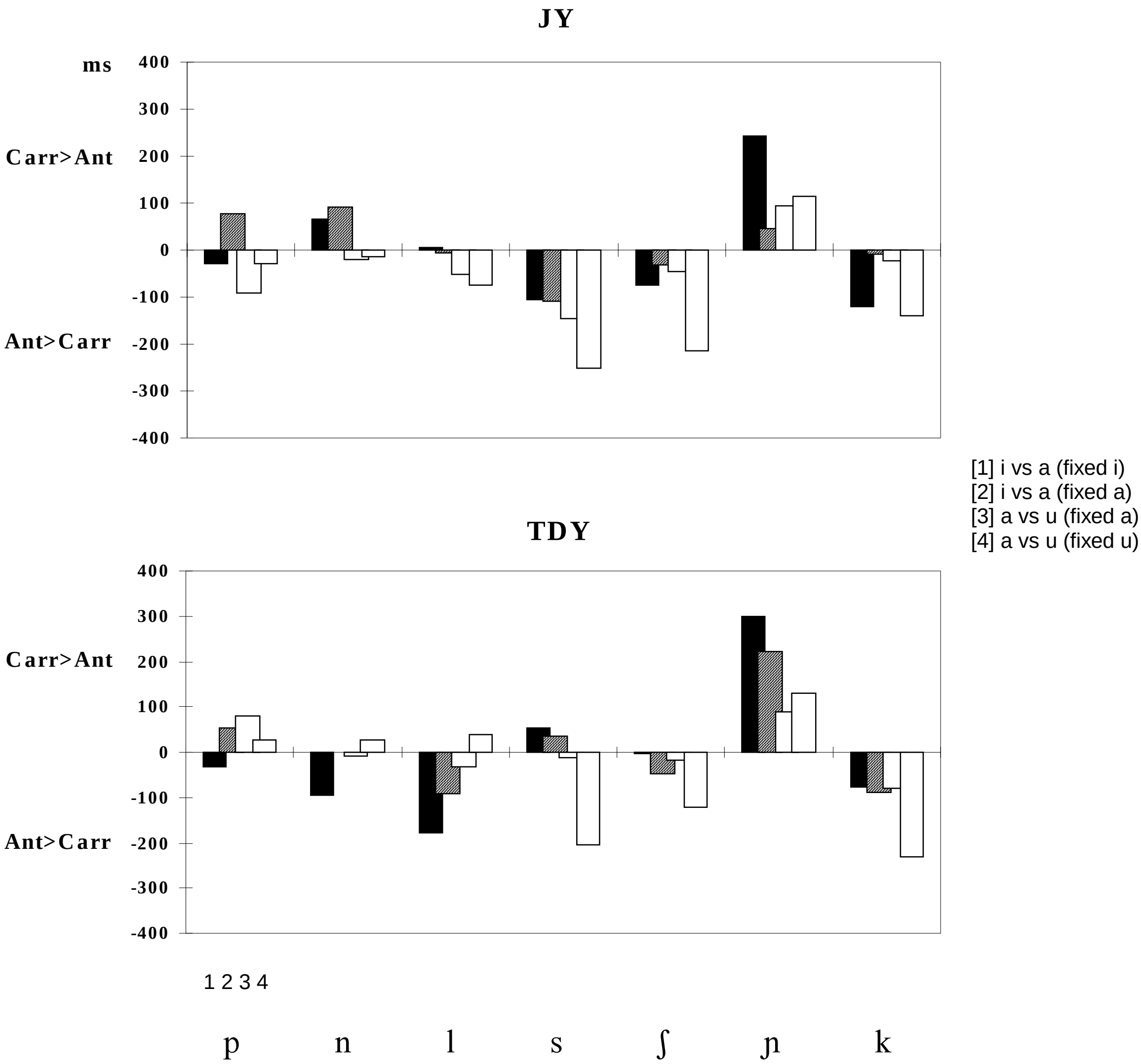


Figure 2. C-A differences in temporal extent of V-to-V coarticulation for JY (top) and TDY (bottom) across speakers. Data are displayed for different consonants and changing and fixed vowels. Positive values indicate prevalence of the carryover over the anticipatory direction, and negative values the opposite relationship.

Jaw and tongue dorsum coarticulation data for /p/, /n/ and dark /l/ do not exhibit a common directionality pattern. Overall the anticipatory component appears to be more salient for TDY than for JY in /VlV/ sequences when /i/ is involved.

4 Discussion

In agreement with predictions from the DAC model, directionality patterns in JY coarticulation were found to depend on specific articulatory requirements for the production of consonants. Dorsals /ɲ/ and /k/ in all sequences and the lingual fricatives /s/ and /ʃ/ in sequences without /i/ exhibit similar trends in coarticulatory direction for TDY and JY, and high TDY-JY correlations in coarticulatory duration. Indeed, /ɲ/ favors the carryover component and /s/, /ʃ/ and /k/ the anticipatory component, while both prevalent coarticulatory directions yield specially long V-to-V effects in most cases. Coincident trends in coarticulatory direction for tongue dorsum and jaw along the vertical dimension in VCV sequences with dorsal consonants may result from the strong muscle linkages between these two articulatory structures: the jaw assists the tongue dorsum during the production of dorsal consonants, and overshoot in JY activity may take place in VCV sequences with two or three dorsal segmental units. Manner requirements may help explain why lingual fricatives favor anticipation for both TDY and JY in sequences without /i/. Prevalence of anticipation over carryover for velars may be related to forward tongue dorsum motion during the velar closure period causing prominent C-to-V anticipation to occur.

This cooperative action between the jaw and the tongue dorsum was found to be less effective for consonants produced with the tongue front and a lower jaw. Indeed, /l/ and /n/ show differences in coarticulatory direction and duration between TDY and JY in sequences with /i/, namely, longer anticipatory effects and shorter carryover effects for TDY than for JY.

In Recasens (2002) vocalic effects were found to be generally longer in the context of fixed back /a/, /u/ vs front /i/ (TDX, TDY) and of fixed low back /a/ vs high back /u/ (TDX, TDY, TTX). However, articulatory overshoot may cause dorsal consonants to exhibit long tongue dorsum effects to occur during fixed /i/. In the present study, JY effects were also found to be longer during fixed /i/ than during fixed /a/ not only in VCV sequences with dorsal consonants (and longer during fixed /a/ than during fixed /u/, as expected). This finding could be associated with the relatively slow motion of the mandible in speech.

Data presented in this paper indicate that coarticulatory effects for different articulators may be strongly related to patterns of interarticulatory coordination. Thus, the jaw and the tongue dorsum may show an analogous coarticulatory behavior for consonants exhibiting a close interaction between the two articulatory structures (e.g., dorsals) but a different coarticulatory behavior for those other consonants for the production of which the tongue dorsum and the jaw do not act coordinatively (e.g., dentoalveolars). To a large extent, certain coarticulation aspects (e.g., prevalence of carryover over anticipation for V-to-V effects on fixed /i/) are jaw specific and may be associated with the massive structure of the mandible in speech.

In contrast with F1 coarticulation data (Recasens and Pallarès, 2001), V-to-V effects in jaw vertical displacement were not always found to be longer at the carryover vs anticipatory level in fixed /a/ condition. Indeed, mandibular effects for fricatives and, to some extent, for /p/ and /l/ on /a/ were reported to favor the anticipatory vs carryover component. Regarding the fixed /i/ condition, the JY data are in accordance with the F1 data in favoring anticipation for /s/ and /ʃ/ but in disagreement with them in not favoring the carryover direction for /l/ and velars. Other dorsals, i.e., alveolopalatals, appear to favor the carryover component however. Differences between the results reported in our F1 study and in the present JY investigation could be partly due to the fact that isolated VCV sequences were used in the former vs the latter. They may also reflect the possibility that F1 coarticulatory effects are related to other articulatory factors besides jaw vertical displacement.

Acknowledgments

This work was supported by projects DGES PB96-0103 and BFF2000-0075-C02-01 of the Ministry of Education and Science of Spain and by research group 2001SGR00425 of the Government of Catalonia. I would like to thank Maria Dolors Pallarès for help with the statistical analysis.

References

Fletcher, J. and Harrington, J. (1999) Lip and jaw coarticulation, in *Coarticulation. Theory, data and techniques*, edited by W.J. Hardcastle and N. Hewlett (Cambridge University Press, Cambridge), pp. 164-178.

Keating, P.A., Lindblom, B., Lubker, J., and Kreiman, J. (1994) Variability in jaw height for segments in English and Swedish VCVs, *Journal of Phonetics*, 22, 407-422.

Magen, H.S. (1997) The extent of vowel-to-vowel coarticulation in English, *Journal of Phonetics,* 25, 187-206.

Recasens, D. (2002) An EMA study of VCV coarticulatory direction, *Journal of the Acoustical Society of America*, 111.

Recasens, D. and Pallarès, M.D. (2001) Coarticulation, blending and assimilation in Catalan consonant clusters, *Journal of Phonetics*, 29, 273-301.

Recasens, D. and Pallarès, M.D. (2000) A study of F1 coarticulation in VCV sequences, *Journal of Speech and Hearing Research*, 43, 501-512.

Recasens, D., Pallarès, M.D. and Fontdevila, J. (1997) A model of lingual coarticulation based on articulatory constraints, *Journal of the Acoustical Society of America*, 102, 544-561.

Sussman, H.M., MacNeilage, P.F., and Hanson, R.J. (1973) Labial and mandibular dynamics during the production of bilabial consonants: Preliminary observations, *Journal of the Acoustical Sociey of America*, 16, 397-420.

Tuller, E., Harris, K.S., and Gross, R. (1981) Electromyographic study of the jaw muscles during speech, *Journal of Phonetics*, 9, 175-188.

Words without Vowels: Phonetic and Phonological Evidence from Tashlhiyt Berber[1]

Rachid Ridouane
UPRESA 7018 CNRS-Université Paris 3

Abstract

This article deals with the Tashlhiyt dialect of Berber (henceforth TB) spoken in the southern part of Morocco. In TB, words may consist entirely of consonants without vowels and sometimes of only voiceless obstruents, e.g. *tftχtstt* "you rolled it (fem)". In this study we have carried out acoustic, video-endoscopic and phonological analyses to answer the following question: is schwa, which may function as syllabic, a segment at the level of phonetic representations in TB? Video-endoscopic films were made of one male native speaker of TB, producing a list of forms consisting entirely of voiceless obstruents. The same list was produced by 7 male native speakers of TB for the acoustic analysis. The phonological analysis is based on the behaviour of vowels with respect to the phonological rule of assibilation. This study shows the absence of schwa vowels in forms consisting of voiceless obstruents.

1 Introduction

Berber is spoken in large parts of North Africa and especially in Morocco and Algeria. Berber is not a standard or a codified language; it exists only through its dialectal or regional realizations. Tashlhiyt Berber (henceforth TB), which is spoken in the Southern part of Morocco, is a homogeneous dialect which enables all the Chleuhs to communicate without difficulties. This apparent homogeneity, however, conceals a certain amount of heterogeneity. Following the work of Boukous (1994), TB may be subdivided into three subsystems :

- The occlusive subsystem spoken in Agadir and its suburbs. It is considered to be the "received Tashlhiyt pronunciation".
- The fricative subsystem spoken in Haha which spirantizes the non coronal obstruents /b, k, g/.
- The sibilant subsystem spoken in the Anti-Atlas mountains where /t/ and /d/ are realized as [s] and [z] respectively.[2]

At the phonological level, TB varieties share the same phonological system founded upon the same correlations. The vowels of TB are /a, i, u/. In this article, all transcriptions not enclosed between slanted lines are phonetic representations given in a broad transcription. The phonetic transcription of a form is meant to represent its pronunciation in isolation. The schwa vowel, which is the main topic of this paper, is transcribed as [e]. Complex consonant clusters are a common occurrence in all the varieties of TB at the underlying level as well as at the phonetic one. Indeed, widely attested forms may consist entirely of consonants without vowels and sometimes of only voiceless consonants, e.g. *tqssf* "it shrunk (fem)", *tfktstt* "you gave it (fem)". How should these words be syllabified and, being deprived of full vowels, sonorants and even of voiced obstruents, which segments may function as syllabic?

---

[1] I am grateful to A. Rialland, F. Dell, N. Clements and J. Vaissière for discussions and comments on an earlier version of this paper. I am also grateful to Dr. Lise Crevier for her comments and assistance for collecting fiberscopic data at Laenec Hospital, Paris. I would also like to thank B. Gautheron and C. Zeroual for technical assistance and all the subjects for their kind participation in this study. Errors are ours only.

[2] Each subsystem is represented in this paper by the area where it is spoken: Agadir for the occlusive, Haha for the spirant and Anti-Atlas for the sibilant.

These questions have aroused much interest among phonologists and phoneticians (Elmedlaoui, 1985, Dell and Elmedlaoui, 1985, 1988, 1996a-b, 1997; Louali and Puech, 1996, 1999a., 1999b.; Boukous, 1990; Coleman, 1996, 1999; Clements, 1997). Dell and Elmedlaoui claim that surface syllables may consist only of consonants with CC and CCC (where the second C is the nucleus) existing along side the conventional syllable types CV and CVC. The syllable organization is determined on the basis of local relations in sonority. According to Dell and Elmedlaoui, there exist ultra short vowels but they occur only adjacent to voiced consonants. No voiced vocoid, however short it is, can be heard in sequences of voiceless obstruents. Dell and Elmedlaoui's data are based on the variety of Tashlhiyt spoken in Imdlawen valley.

Coleman (1996, 1999) argues for a different claim. According to him, epenthetic vowels exist in TB and they are the phonetic realizations of syllable nuclei. Where no vowel epenthetic is evident, it can be regarded as hidden by the following consonant. In the framework of the co-production model, Coleman interprets the syllabic consonants as the co-produced realization of phonological vowel and consonant. Coleman's analysis is based on published material and on his own work on the variety of TB spoken in Agadir (Coleman 1996) and the variety of TB spoken in Imdlawen valley (Coleman 1999).

Louali and Puech, basing their arguments on some acoustic and perceptual analyses, conclude that a vocoid is always present in the realization of a TB word, if not a full vowel then a schwa vocoid, even within words composed only of unvoiced obstruents. Louali and Puech's analyses are based on their own work on different varieties of TB.

1.1 Aim

If, as Dell and Elmedlaoui claim, schwa is only an aspect of the realization of a voiced consonant, and not a segment, then one should not find schwa and voicing or any vocalic gesture in a word consisting of voiceless obstruents. We don't know of a mechanism that would introduce voicing in a sequence deprived of [+ voiced] segments. On the other hand, if these words contain schwas, then that would mean that the phonetic representation of TB contains at least four vowels: the realizations of /a, i, u/ and one segment schwa which can thus occupy the nucleus of a syllable. We have investigated the three varieties of TB. Our focus in this paper is the following: to what extent can we speak of voiceless obstruents as being syllabic nuclei? One way of answering this question is to determine if schwa is a segment at the level of phonetic representations of TB. We will examine through spectrography and video-endoscopy the presence or the absence of voicing or vowel gestures in sequences of voiceless obstruent clusters. A phonological argumentation will also be developed.

2 Spectrographic analysis

2.1 Method

For the spectrographic analysis, the 23 forms listed in (1), consisting of widely attested verbs in TB, were produced by 7 male adult TB native speakers. These verbs only contain voiceless sequences:

(1)

Sequences of two voiceless obstruents :

ks	"feed on"
fk	"give"

Sequences of three voiceless obstruents :

kst	"feed it on"
kt	"give it"
kks	"take off"
sχf	"fade away"
ftħ	"operate"

Sequences of four voiceless obstruents :

tftħ	"she operated"
tkʃf	"it is dirty"
tfss	"she is quiet"
fqqs	"irritate"

Sequences of five voiceless obstruents :

tʃtft	"you crushed"
tkkst	"you took off"
!tχtft[3]	"you stole"
kkstt	"take it off (fem)"
tfsχt	"you cancelled"
tqssf	"it shrunk (fem)"

Sequences of six voiceless obstruents and more :

sfqqst	"irritate him"
tfktstt	"you gave it"
tsskʃft	"you made it dirty"
tkkststt	"you took it off (fem)"
tftktstt	"you sprained it (fem)"
tftχtstt	"you rolled it (fem)"

We have taken into consideration the different varieties of TB and chosen our subjects according to their native TB subsystem. This breakdown is illustrated in Table 1 below:

Table 1. The breakdown of the subjects according to their native TB subsystem :

Subject	Subsystem	Age
E	Agadir	31
H	Agadir	26
S	Anti-Atlas	33
B	Anti-Atlas	30
K	Haha	63
A	Haha	28
R	Haha	30

[3] "!" indicates that all the segments of the word are dorsopharyngealized.

Each subject produced each form in (1) at least three times. Only the first three utterances of each form were analysed using Unice software. We looked into the acoustic realizations of the items (1) to determine the presence or the absence of schwa vowels.

2.2 Results and discussion

A first result of our spectrographic data shows the predominance of the realizations of forms (1) with no voiced vocoids. Only 68 out of 465 utterances are realized with schwa vowels. Table 2 below summarizes these different realizations for each subject.

Table 2. The realizations of schwa vowel in the utterances of 7 subjects.

	Agadir		AA		Haha			
	E	*H*	*S*	*B*	*K*	*A*	*R*	**TOTAL**
Fk	3	3	0	0	0	0	0	**6/21**
Ks	3	3	0	0	0	0	0	**6/21**
Fkt	3	3	0	0	0	0	0	**6/21**
Kst	3	3	0	0	0	0	0	**6/21**
Ftħ	3	0	0	0	0	0	0	**3/21**
Kks	2	0	0	0	0	0	0	**2/21**
sχf	3	0	0	0	0	0	0	**3/21**
Fqqs	0	0	0	0	0	0	0	**0/21**
Tfss	3	0	0	0	0	0	0	**3/21**
Tftħ	3	0	0	0	0	0	0	**3/21**
tkʃf	3	0	1	0	0	0	0	**4/21**
Tkkst	3	0	0	0	0	0	0	**3/21**
Kkstt	3	0	0	0	0	0	0	**3/21**
tfsχt	3	0	0	0	0	0	0	**3/21**
tʃtft	3	0	0	0	0	0	0	**3/21**
Tqssf	3	0	0	0	0	0	0	**3/21**
!tχtft	3	0	0	0	0	0	0	**3/21**
sfqqst	3	0	0	1	0	0	0	**4/21**
tsskʃft	3	0	0	0	0	0	0	**3/21**
Tfktstt	1	-[4]	0	0	0	0	0	**1/18**
tftktstt	-	-	0	0	0	0	0	**0/15**
tftχtstt	0	-	0	0	0	0	0	**0/18**
tkkststt	-	-	0	0	0	0	0	**0/15**
								68/465

The number of the realizations with a voiced vocoid is indicated in the relevant box. Subject B, for instance, did not pronounce any schwa during his three realizations of the form *fk* (0), whereas subject E realised the same form with a schwa during his three utterances (3).

The three subjects of Haha (K, A and R) never pronounced schwas within sequences of voiceless obstruents. For the Anti-Atlas (AA) subjects only two utterances contained schwas at the final position in the realizations of: *tkʃfe* for S and *sfqqste* for B. By contrast, voiced vocoids are widely attested in the realizations of Agadir subjects: H and especially E. How should we explain this variation? Do these differences reflect a difference in the syllabic

[4] This symbol means that the subject considers his three utterances to be incorrect.

structure between the different TB subsystems? Or do they reflect only individual variation? Before answering these questions, let us determine the distribution of these schwas in the realizations of E and H. Table 3 below sums up these distributions :

Table 3. The distribution of schwa vowels in the realizations of subjects E and H.

	E	*H*
Fk	Fk**e**	fk**e**
Ks	Ks**e**	ks**e**
Fkt	Fkt**e**	fkt**e**
Kst	Kst**e**	kst**e**
Ftђ	ftђ**e**	---
Kks	kks**e** 2/3	---
sχf	sχ**e**f	---
tfss	tf**e**ss	---
Tftђ	1/3 tft**e**ђ ; 2/3 tftђ**e**	---
tkʃf	tkʃf**e**	---
tkkst	tkkst**e**	---
kkstt	kkste	---
Tfsχt	tfsχt**e**	---
tʃtft	tʃtft**e**	---
tqssf	tqssf**e**	---
! tχtft	!tχtft**e**	---
sfqqst	1/3 sf**e**qqst ; 2/3 sfqqst**e**	---
tsskʃft	tsskʃft**e**	---
tfktstt	tfktstt**e** 1/3	-

Some important observations may be made from the data which we have presented so far. Firstly, the voiced vocoid is mainly attested at the end of the utterances. Subject H realizes all his schwas at this position. Schwa is attested at the medial position in only four items for subject E. Consider below, the spectrogram of the form *fke* as realized by H :

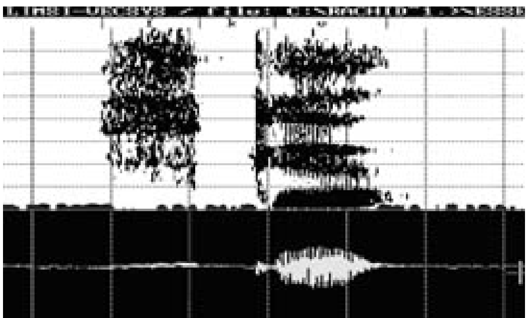


Figure 1. A spectrogram of *fk* « give ! » par H.[5]

[5] The horizontal grid lines in all the spectrograms mark divisions into KHz.

The presence of a schwa vowel in this form was also observed by Louali and Puech (1999a, Figure 2) "*Figure 2 manifests the presence of a schwa vowel after the release of the stop, which then occupies the syllabic position*." (p. 39). Louali and Puech added that: "*No vocoid is attested in the spectrogram in Figure 3* [for their form *fkt*] *where the stop also occupies the syllabic position*"[6] (ibid). The fact that schwa is present in one form and absent in the other is not due to the syllabicity of the velar but to the final position the segment occupies. The velar stop is followed by a schwa vowel in Louali and Puech's Figure 2 because it occupies the final position, a position which it no longer occupies in their form *fkt*. One should wonder why schwa is absent not after /k/ but rather after the final /t/. We will show below that this may be adequately accounted for by the fact that this segment is very unstable. The presence of schwa vowels at the end of some utterances[7], especially after a stop, is attested in other languages. This is the case for instance of some French words (*cap*, *bec* etc.). The presence of this schwa aids the hearer to identify the final consonant, occurring in a naturally weak position, by placing it in a prevocalic position where cues to its place of articulation (burst, noise transient, formant transitions) will be present.[8]

A second important observation is that the distribution of schwas may be particularly unstable even within one same subject. Consider for instance the two utterances of the form *sfqqst* as realized by E in a period of less than three seconds. As is shown in the Figure 2, a voiced vocoid is realized between /f/ and /q/ in the first utterance but word finally in the second. Not all the forms realized by E contain schwas. He may realize long voiceless sequences with no vowel at all. This is the case for instance for the items *fqqs, tfktstt* and *tftχtstt*. Below in Figure 3 we see a spectrogram of a sequence of 8 voiceless obstruents realized by E with no schwa vowel.

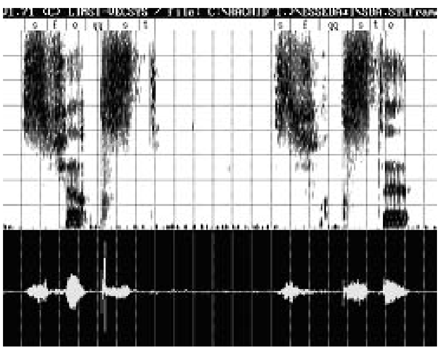


Figure 2. A Spectrogram of two pronunciations of *sfqqst* « irritate him» by E .

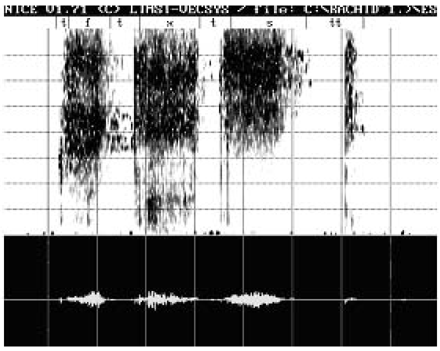


Figure 3. A spectrogram of *tftχtstt* « you rolled it » by E.

An additional important observation is that schwas occur mainly in the Agadir sub-system subjects. Does this mean that the syllabic system of this variety is different from that of the other varieties? This question merits thinking about especially since Coleman (1996) is based on some data uttered by a subject belonging to this area. My own judgement as a native speaker of Haha variety, having lived for a long time in Agadir, prompts me to consider this variations to be mainly due to the influence of Moroccan Arabic (henceforth MA) pronunciation. It is a known fact that TB speakers raised in Agadir or in any other Arabic-

[6] We have translated these two sentences from French.
[7] Compare the case of this schwa with the schwa in the Ath Sidhar Berber dialect spoken in the North of Morocco which never occurs word-finally. In this dialect, an epenthetic schwa, unlike in TB, is inserted in order to syllabify consonants which would otherwise remain unsyllabified (see Dell and Tangi 1992)
[8] Personal communication of N. Clements.

speaking city mimic the MA pronunciation considered to be more prestigious.[9] Subject E, who lived in Agadir, may insert schwas to break up consonant clusters, but this is not acceptable for subject K, for example, who lived in a remote Berber village for more than thirty years before coming to France. This phenomenon is illustrated by the form *sχef*, which is a MA loan word totally integrated in the TB lexicon. Subject E mimicked the MA pronunciation of this form with a schwa that must be pronounced in MA before the last consonant when realized in isolation. Consider the spectrogram of this item as realized by subject E and compare it with the spectrogram of the same form realized by a MA native speaker:[10]

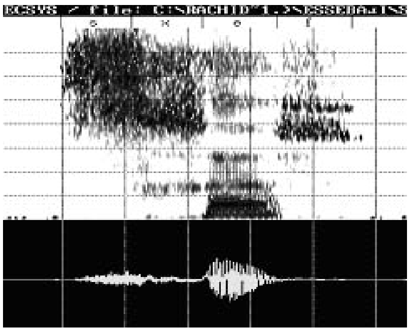


Figure 4. A spectrogram of *sχef* « fade away » by E.

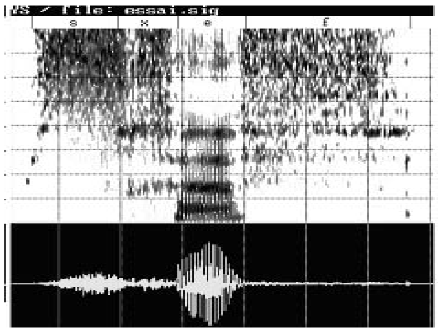


Figure 5. A spectrogram of *sxef* « fade away » by a MA native speaker.

One last observation concerns voiceless schwas. Coleman's and Louali and Puech's claims don't extend to these segments. None of our 7 subjects produces voiceless schwas. Our speakers either pronounce voiced schwas or no schwas at all. A voiceless schwa is a vowel with formants but which lacks the voicing bar. No such segment was detected through our spectrograms. The phonological argument which will be developed below also precludes the presence of this segment.

To summarize, the acoustic analysis allows us to argue for the absence of schwa vowels in voiceless obstruent sequences. But this is not without exceptions. What we have shown in this section is that these exceptions fall into two classes. The first class of exceptions concerns the presence of schwa vowels mainly in the final position of the utterance. The presence of schwa in the periphery is considered to be a cue to identify the final consonant by placing it in a prevocalic position. The second class of exceptions involves the internal schwas, whose presence is due to the influence of MA pronunciation.

9 See Dell and Elmedlaoui (in press), who have observed MA like schwas in the speech of the Chleuhs living in an Arabic-speaking city. Giving several examples for illustration, they observed that the command of TB tends to deteriorate and that adaptations from the dominant language are increasingly resorted to.

10 The two other forms containing a schwa at the internal position in *tfth* and *sfqqst* are all MA loan words. The other form with internal schwa *tfss* is a native TB verb that may be realized in two ways: *tffis* with a full vowel or a vowel-less *tfss*. These two realizations are in free variation.

3 Physiological arguments

3.1 Method

While the configuration of the glottis for voiceless consonants – especially voiceless aspirated stops – in CV, VCV and VC are well accounted for, much work is still to be done for an understanding of laryngeal movements in more complex environments. TB, which allows long voiceless sequences, provides one of the most suitable fields for examining these laryngeal movements.

The most convenient source of information on laryngeal movements in obstruent sequences is a series of articles published by Löfqvist and colleagues. In these articles, sequences of voiceless sounds in American English, Dutch, Swedish, Icelandic and Japanese were studied. Concerning Berber, this paper is the first attempt to deal with such topic in this language.

One native speaker of Haha sub-variety of TB served as subject for this experiment, a 30-year-old male.[11] The video-endoscopic experiment was performed by means of a flexible nasofibroptic laryngoscopy (Olympus ENF P3) with video recording (25 frames/s).[12] A fiberscope was inserted through the nostril of the subject. A camera Sony (XC–999 P) was fixed on the external side of the fiberscope which enabled us to record a video film on a U-Matic Sony tape recorder (VO-5800 PS). The internal side was stabilized a bit over the larynx which provided an immediate visualization of the dynamic behaviour of the laryngeal region. The laryngeal evaluation included the abduction and adduction movements of the vocal folds as well as inward and forward movements of the arytenoid cartilages. A synchronization signal was recorded on one channel of the tape recorder for frame identification. The film was analysed by means of a PC using the programs Adobe Premiere and Adobe Photoshop.

The corpus consisted of the same forms as for the spectrographic analysis (see (1) above). A supplementary data below was also recorded in order to have a view of the dynamic behaviour of the larynx during the production of some French and MA words. The aim is to compare the behaviour of the glottis during the production of French voiceless sequences (underlined sequences in (2) below) and MA words with the state of the glottis during the realization of TB items:[13]

(2) French items

ʃa<u>kst</u>ilo	"every pen"
feli<u>ksf</u>ɔR	"Felix Faur"

(3) MA Items

fseχ	"He cancelled"
fqqes	"irritate"

Each item of (1), (2) and (3) was produced six times in isolation. The rest interval between consecutive items was approximately 2-3 seconds. Our data consists of 162 utterances.

---

[11] I served as a subject for this experiment. My wish was to have at least three subjects, each representing one variety of TB, but all the persons I contacted showed little motivation to take part in the experiment. This is certainly due to the nature of the experiment, which may seem somehow repugnant.

[12] These recordings were made at Laennec Hospital in Paris, with the cooperation of Doctor Lise Crevier.

[13] Notice that the sequence *<u>kst</u>* in *ʃakstilo* is identical to the TB item *kst* "feed it on" and that *fqqes* exists also in TB but with no schwa vowel. This item originates from MA.

3.2 Results and discussion

The difference between vowel-less obstruent sequences and sequences with a schwa vowel can be easily observed by means of video-endoscopy. The two figures below give an immediate visualisation of one state of the glottis during the production of a voiceless obstruent sequence (Figure 6) and one state of the glottis during the production of *fseχ*, a MA item having a schwa vowel before the final consonant (Figure 7).

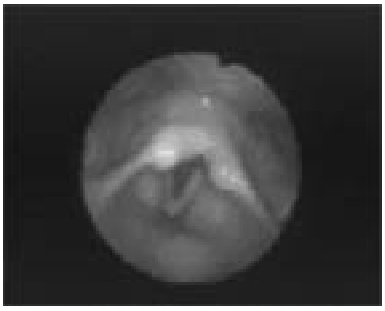


Figure 6. One state of the glottis for TB *[tss**k**ʃft].*[14]

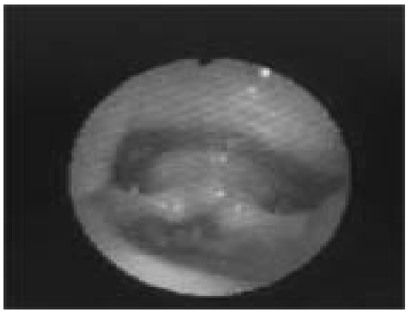


Figure 7. One state of the glottis for MA *[fs**e**χ]*.

We have observed very similar laryngeal movements during the production of items (1). At the initial position, the glottis is widely open. This configuration is identical to the one in Figure 6 above. As is shown in Figure 12[15], for the form *tfsχt*, for example, the state of the glottis for the closing (or onset) phase of the oral stop /t/, which corresponds to the first image, can be described as having an abducted glottis, abducted vocal folds and abducted arytenoid cartilages. The glottis maintains this quasi static position of opening during the period of the utterance of the whole cluster. This configuration of the glottis is the same for all the utterances of the 23 clusters of voiceless obstruents we have analysed, regardless of the number of consonants they have (see Figure 14 for *tftktstt,* a sequence of 8 voiceless obstruents). All these forms show an uninterrupted devoicing through the sequence. When the arytenoids are this far apart the vocal folds can not be set into vibration.

We have also observed a general tendency to lower the glottis towards the end of each statement. There is also a general tendency to bring the vocal folds closer, but still not adducted, at the end of each form. The arytenoid cartilages remain separated. This tendency is not due to the nature of the final consonants since it is attested after stops as well as after fricative segments as is shown in Figures 12 and 13. A parallel may be drawn between this observation and the fact that certain subjects may realize schwas at the end of the utterances. We might consider that this tendency may attend a further degree of glottal closing to reach a vocalization phase for certain subjects as is the case for H, E, B and S.

As outlined above, we have examined some French voiceless sequences and compared them with TB items. This experiment was designed to investigate further evidences proving the absence of vowel gestures in TB voiceless sequences. Another purpose was to provide some more observations of the laryngeal movements in the two languages.

The French sequence /k#st/ shows an interrupted devoicing through the sequence. The same observation showing separation of the vocal folds occurs in TB *kst*. Below, two states of

[14] If a single segment is underlined, as is the case in Figure 6, the image shows a frame corresponding to a medial portion of that segment, and if two segments are underlined, it shows a frame at the transition between the two.

[15] The Figures 12 to 19 are in the Index. The primes in these figures indicate the closing phase as in [‘t] and the release phase as in [t’].

the glottis during the production of the two sequences. Both images correspond to the onset phase of the final stop /t/ (see Figures 15 and 16 for a visualization of the whole sequences).

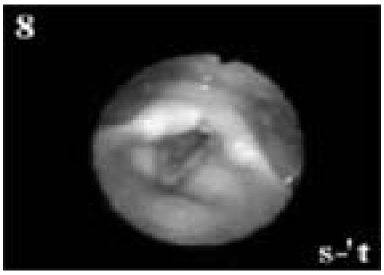


Figure 8. One state of the glottis during the production of TB *[k**st**]*.

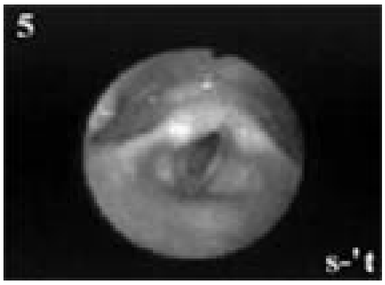


Figure 9. One state of the glottis during the production of French *[ʃak**st**ilo]*.

Differences between the two sequences can be detected. Our observations show that the degree of glottal opening is not the same in the two forms. In TB *kst* the glottis opening is wide and maintains this degree through the whole sequence.[16] In French /k#st/ however, the glottis is slightly open during the onset and the closure phases of the stop /k/ and opens gradually to reach its maximal width at the onset phase of the final /t/ then starts closing gradually.

As we have pointed out earlier, clusters of voiceless consonants provide one of the most suitable fields for examining the processes of coarticulation and coproduction at the laryngeal level by examining how the simple ballistic-looking pattern of abduction-adduction found in singular consonants followed by a vowel is modified when sequences of voiceless consonants occur. One important issue to examine is the relative extent of anticipatory effects.

In analysing sequences of voiceless obstruents followed by a vowel such as French */k#st/* in *ʃakstilo*, we have observed that the glottis moves gradually to a closing phase during the realization of the dental /t/ which is followed by a vowel. We believe that this an anticipating effect of the following vowel. The glottis seems to anticipate the closing gesture necessary for the realization of a vowel. We have observed the same phenomenon in TB sequences of voiceless obstruents followed by a vowel such as *tsskti* "she reminded" in Figure 19.[17] As we shall see below, the same anticipating effect is observed in MA items as well.

So far different arguments have been developed showing the absence of vowel gestures in TB voiceless items. One last argument is provided by the comparison of the two realizations of the forms *fqqs*. As is outlined in the acoustic section, some TB items are MA loan words. *fqqs* is one such item. In MA, an epenthetic schwa is inserted before the last consonant. Such vowel does not exist in TB. Figure10 below shows a clear vowel gesture and thus voicing after the occlusion of the geminate stop /qq/ in the MA *fqqes*. Whereas in Figure 11, no such vocalic gesture is attested either in this position or elsewhere in the sequence (see Figures 17 and 18 in the Index for the whole sequences of the two forms).

[16] The observation deduced from this material suggests a slightly different picture from the observations of Löfqvist and Yoshioka (1980) who argued that: "There is little, if any evidence that the glottis ever opens and maintains a static position in speech."

[17] The fact that the glottis anticipates the realization of the following vowel is more obvious when we compare the final /t/ in the two Figures 14 and 19.

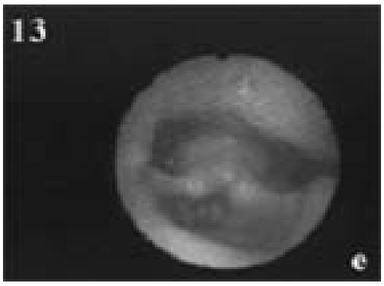


Figure 10. One state of the glottis during the production of MA *[fqqes].*

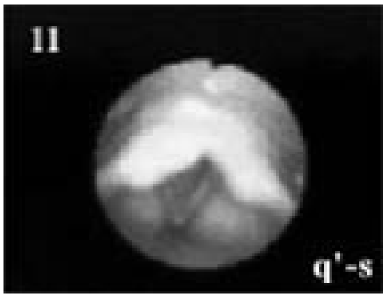


Figure 11. One state of the glottis during the production of TB *[fqqs].*

The anticipatory effect observed in French and TB items is also attested in MA. In fact, after the maximal opening of the glottis maintained from the beginning of the utterance until the onset phase of the geminate /qq/, the glottis closes gradually to reach a maximal closing after the occlusion of the stop.

In French /k#st/, the velar stop is preceded by a vowel as in *fakstilo*. In the MA item *fqqes*, the fricative /s/ is also preceded by a vowel. The laryngeal movements are not the same for the two segments. In MA, we observed that the onset of glottal opening is very early with a maximal opening immediately after the vowel (see Figure 18). In French, the opening is rather slow and the glottis opens gradually before reaching a maximal opening after at the onset phase of the final /t/. One possible reason for these differences is due to the nature of the two consonants following the vowel: a stop /k/ in the French item and a fricative /s/ in MA.[18] In fact, there is a fairly widespread finding in literature that abduction is more rigorous in fricatives than in stops. Our observations are consistent with the observations of Löfqvist and Yoshioka (1980) and Yoshioka et al. (1981) using respectively Swedish and American English as well as with other literature findings[19]. Löfqvist and McGarr (1987: 399) (cited in Hoole 1997) discuss reasons for the larger glottal gestures in fricatives: “The larger gesture for a voiceless fricative is most likely due to the aerodynamics of fricative production, in that a large glottal opening not only prevents voicing but also reduces laryngeal resistance to air flow and assists in the build up of oral pressure necessary for driving the noise source.”

To summarize, our physiological observations show the absence of vocalic gestures in all the utterances of the 23 forms analysed. All these forms show separation of the vocal folds and the arytenoid cartilages. Just as there are no vowel gestures in French voiceless sequences, so is the case in TB. The particular laryngeal configuration in French sequences is believed to be mainly due to anticipatory effects. The comparison with MA items shows a clear difference between a sequence having a schwa vowel and a sequence deprived of any vocalic gesture. Various observations either in accord or in disaccord with some literature findings are also deduced. These observations will be dealt with in some detail in forthcoming studies.

4 Phonological argumentation

The phonological analysis is based on the behaviour of vowels with respect to the phonological rule of assibilation according to which the dental stops /t/ and /d/ become /s/ and /z/ respectively. The dental stops maintain their occlusion when they are immediately in contact with a coronal consonant e.g. /itri/ → *itri* “star”. When the dentals are separated from the coronals by one of the vowels /a, i, or u/, the process of assibilation operates e.g. /tura/ → *sura* “she wrote”. The aim of the phonological analysis is to check the behaviour of dental

[18] The maximal glottal opening tends to coincide with the location of [s] in French as well (see Figure 16).
[19] See Hoole (1997) for a summary of this literature and the references therein.

stops when separated from other coronal consonants, according to certain analyses and transcriptions, by the vowel [e]. Our analysis is based on the data analysed by Louali-Raynal (1999). We have chosen this data on purpose so as to compare easily her conclusions with the conclusions the same author published in another work (Louali and Puech 1999b) on syllabification in the same variety of TB.

A general claim is that schwa is invisible in regard to all relevant phonological processes and constraints in TB. For example, it is invisible in regard to assibilation. Assibilation is a phonological process attested in the Anti-Atlas variety whereby dentals /t/ and /d/ are realized as sibilants [s] and [z] respectively:

(4)	/tifawt/	sifaws	light
	/tabawt/	sabaws	bean
	/tafukt/	safuks	honey
	/tilkit/	silkis	louse
	/tirgit/	sirgis	embers
	/tismegt/	sismegs	slave

The dentals maintain their occlusion when they are in contact with another coronal segment:

(5)	/adrar/	adrar	mountain
	/itri/	itri	star

The maintain of occlusion does not operate across the vowels /a, i, u/:

(6)	/tura/	sura	she wrote
	/idurar/	izurar	mountains
	/issuda/	issuza	she rode
	/tarwa/	sarwa	children
	/tirγi/	sirγi	heat

Louali-Raynal (1999) provided the verb below for illustration. This form according to her transcription contains a schwa vowel[20]:

(7)		terkez	she wore

Let us suppose that this vowel exists and that /trkz/ contains a schwa between the coronal clusters /t/ and /r/. The schwa here should normally make assibilation possible since /t/ and /r/ are not adjacent. If assibilation took place in (7), it would be pronounced in a way that should be represented as *serkez* in the transcription of Louali-Raynal. But this is not the case, the dental /t/ maintains its occlusion. If we admit the presence of a vowel [e] in this form, we should then explain why vowels /a, i, u/ don’t obstruct assibilation as in (6) while the vowel [e] does. This problem does not appear in our analysis in which [e] is not a segment of TB.[21]

The comparison with some Berber dialects which contain a vowel [e] is very instructive. The Beni Iznassen Berber dialect spoken in the North of Morocco spirantizes the

[20] In many works on Tashlhiyt and not only in Loauli-Raynal (1999), consonant clusters in this environment are as a rule separated by a schwa.
[21] The argumentation developed here concerning assibilation is analogous in its principle to that of F. Dell and M. Elmedloui (2000). Their argumentation makes use of regressive devoicing.

dental /t/ except when preceded immediately by /m/[22] as is shown by the data below (from El Kirat 1987: 216):

(8)	θandamt	regret
	θħumt	ditrict

When the schwa separates the labial /m/ from the dental, spirantization takes place:

(9)	lχeðmeθ	work
	lħekmeħ	magical power.

Unlike in TB, the surface representations of Beni Iznassen Berber contain a vowel [e] which, by separating the dental /t/ from the labial /m/, makes spirantization possible[23].

5 Conclusion

The different arguments we have developed so far prove that schwa is not a segment at the level of phonetic representations in TB. The acoustic arguments showed the predominance of realizations of long sequences of voiceless obstruents with no schwa vowel. The presence of this vowel in the final position of some utterances is believed to be a cue used for the hearer to identify the final consonants by placing them in a prevocalic position. Its presence in the internal position is due to the influence of MA. The physiological arguments exclude the presence of vocalic gestures in these sequences, where the vocal folds and the arytenoid cartilages remain widely separated and the glottis clearly open. The possibility of the presence of a voiceless schwa is excluded by our phonological analysis. No vowel, be it voiceless, separates dentals and coronals in our data. This paper shows that many TB words lack phonological or phonetic vowels. Since no vowel exists in the forms similar to (1), any consonant may then occupy the position of a syllable nucleus, even a voiceless stop. Relying on these conclusions, we are conducting a study to determine if the syllabic consonants have some phonetic properties that differentiate them from their non syllabic counterparts.

[22] Spiratization is inhibited as well when preceded by /l, n/.

[23] The argumentation developed here does not take into account a possible rule of schwa epenthesis which, if it exists, may follow and thus counterfeed the rule of assibilation. On the absence of schwa epenthesis in TB, see Dell and Elmedlaoui (1996).

References

Benguerel, A. & Bhata, T. (1980), “Hindi stop consonants: an acoustical and fiberscopic study”, Phonetica 37, pp. 134-148.

Boukous, M. (1990), “Vocalité, sonorité et syllabicité”, Awal, n° spécial: Hommage à Mouloud Mammeri, pp. 203-218.

Boukous, M. (1994), “Variation phonique et Compétence globale: le cas du parler d’Agadir ”, Publications de la Faculté des Lettres. Rabat.

Clements, N. (1997), “Berber syllabification: Derivations or Constraints ?”, in Derivations and Constraints in Phonology, Iggy Roca eds., Clarendon Press, Oxford, pp. 289-330.

Coleman, J. (1996), “Declarative Syllabification in Tashlhiyt Berber”, in: Current trends in phonology, J. Durand & B. Laks (eds), CNRS, Paris X & University of Salford, University of Salford Publications, pp. 175-216.

Coleman, J. (1999), “The nature of vocoids associated with syllabic consonants in Tashlhiyt Berber”, Proceedings of the 14th International Congress of Phonetic Sciences, San Fransisco, 1-7 August 1999, pp. 735-738;

Dell, F. & Elmedlaoui, M. (1985), “Syllabic Consonants and Syllabification in Imdlawn Tashlhiyt Berber”, Journal of African Languages and Linguistics 7, pp. 105-130.

Dell, F. & Elmedlaoui, M. (1988), “Syllabic consonants in Berber: Some new evidence”, Journal of African Languages and Linguistics 10, pp. 1-17.

Dell, F. & Elmedlaoui, M. (1996), “Nonsyllabic transitional vocoids in Imdlawn Tashlhiyt Berber”, in: Current trends in phonology, J. Durand & B. Laks (eds), CNRS, Paris X & University of Salford, University of Salford Publications, pp. 217-244.

Dell, F. & Elmedlaoui, M. (1997), “La syllabation et les géminées dans la poésie berbère du Maroc (dialecte chleuh)”, in: Cahiers de Grammaire N° 22, pp. 1-95.

Dell, F. & Elmedlaoui, M. (2000), “Syllabification in Moroccan Arabic and how it differs from that in Tashlhiyt Berber”, Proceedings of the 5th International Conference on Afroasiatic Languages, Paris, June 28-30.

Dell, F. & Elmedlaoui, M. Syllables in Tashlhiyt Berber and in Moroccan Arabic. In press.

Dell, F. & Tangi, O. (1993), “Syllabification and Empty nuclei in Ath-Sidhar Rifain Berber”, Journal of African Languages and Linguistics 13, pp. 125-162.

El Kirat, Y. (1987), Spirantization in the Beni Iznassen dialect, diachrony and synchrony. Unpublished thesis of D.E.S, Mohammed V University, Rabat.

Elmedlaoui, M. (1985), Le parler berbère chleuh d’Imdlawn (Maroc); segments et syllabation. Unpublished thèse de Troisième cycle, Université de Paris VIII.

Kagaya, R. (1974), “A fiberscopic and acoustic study of the Korean stops, affricates and Fricatives”, Journal of Phonetics 2, pp. 161-180.

Löfqvist, A. & Yoshioka, H. (1980), “Laryngeal activity in Swedish Obstruent clusters,” The Journal of the Acoustical Society of America, 68, pp. 792-801.

Louali-Raynal, N. (1999) “La spirantisation en berbère” Afroasiatica Tergestina Papers from the 9th Italian Meeting of Afro-Asiatic (Hamito-semitic) Linguistics, Treiste, 23-24 April 1988, Unipress, pp. 271-298.

Louali, N. & Puech, G. (1996), “Syllabic consonants in Tashlhiyt Berber : the case of unvoiced stops”, Communication about : The phonology of the World’s Languages : The Syllable, Pezenas, 21 au 24 juin.

Louali, N. & Puech, G. (1999 a), “Syllabification in Berber, the case of Tashlhiyt”, Proceedings of the 14th International Congress of Phonetic Sciences, San Fransisco, 1-7 August 1999, pp. 747-750.

Louali, N. & Puech, G. (1999 b), “La syllabe en berbère Tachelhit”, in Syllabes: IIèmes Journées d’Etudes Linguistiques, Université de Nanterre, S.Wauquier – Gravelines et S. Ferré, eds.

Harris, G. H. (1999), “States of the glottis for Voiceless Plosives”, Proceedings of the 14th International Congress of Phonetic Sciences, San Fransisco, 1-7 August 1999, pp. 2041-2044.

Hoole, P. (1997), “Laryngeal coarticulation : coarticulatory investigations of the devoicing gesture”, Forschungsberichte des Instituts für Phonetik und Sprachliche Kommunikation der Universität München (FIPKM) 35, pp. 89-99.

Yoshioka, H., Löfqvist, A. & Hirose, H. (1981), “Laryngeal adjustments in the production of consonant clusters and geminates in American English”, The Journal of the Acoustical Society of America, 70, pp. 1615-1623.

The Phonetics and Phonology of some Syllabic Consonants in Southern British English

Zoë Toft
School of Oriental and African Studies
University of London

1 Introduction

This article presents new experimental data on the phonetics of syllabic /l/ and syllabic /n/ in Southern British English and then proposes a new phonological account of their behaviour. Previous analyses (Chomsky and Halle 1968:354, Gimson 1989, Gussmann 1991 and Wells 1995) have proposed that syllabic /l/ and syllabic /n/ should be analysed in a uniform manner. Data presented here, however, shows that syllabic /l/ and syllabic /n/ behave in very different ways, and in light of this, a unitary analysis is not justified. Instead, a proposal is made that syllabic /l/ and syllabic /n/ have different phonological structures, and that these different phonological structures explain their different phonetic behaviours.

This article is organised as follows: First a general background is given to the phenomenon of syllabic consonants both cross linguistically and specifically in Southern British English. In §3 a set of experiments designed to elicit syllabic consonants are described and in §4 the results of these experiments are presented. §5 contains a discussion on data published by earlier authors concerning syllabic consonants in English. In §6 a theoretical phonological framework is set out, and in §7 the results of the experiments are analysed in the light of this framework. In the concluding section, some outstanding issues are addressed and several areas for further research are suggested.

2 Background

In order to discuss syllabic consonants we must clarify what we mean by the term. First we must establish what is meant by 'syllabic', or indeed, 'syllable'. In trying to establish what indeed a syllable is, many phonologists have attempted to describe and explain the possible forms a syllable may take. Typically a syllable is described as consisting of a vocalic centre, optionally accompanied by a consonantal onset or coda, either of which may be complex (cf. Selkirk 1982/1999:329, Blevins 1995:216). In most languages, every syllable has a vowel at its centre. However, some languages allow segments which are not traditionally classed as vocalic to form the nucleus of a syllable, for example the /n/ in /bʌtn/ 'button' or the /l/ in /mʌdl/ 'muddle', and it is these segments which are known as syllabic consonants.

One survey (Bell 1978) cites 85 languages with syllabic consonants yet relatively little work has focused on their phonological or phonetic nature. Are syllabic consonants phonetically different from their non-syllabic counterparts? Can syllabic consonants be flanked by (complex) onsets or codas in the same way vocalic nuclei can be? What contextual restrictions are there on syllabic consonants? Answers to such questions will help refine our understanding of what it means to be a syllable.

(1) Examples of so called syllabic consonants

[ejprn̩]	'apron'	English	Bloomfield 1933:122
[st̩r̩]	'stirrer'	English	Bloomfield 1933:122
[haːbm̩] or [haːbən]	‘to have’	German	Clark &Yallop 1995: 68
[kikŋ] or [kikən]	‘to look’	NE Dutch varieties	
[m̩tu]	‘man’	Swahili	Clark & Yallop 1995:68
[ń̩lá]	‘big’	Yoruba	Clark & Yallop 1995:68
[br̩no]	‘place name’	Czech	Roach 2000:75
[t̩ɬ̩p^{h}tχ]	‘cut it’	Bella Coola	Laver 1994:240
[t^{h}àq^{h}ɑkálk̩^{h}t̩^{h}]	‘to close a door’	Nex Perce	Laver 1994:240
[ʔm̩ː^{3} m̩ʔ^{4}]	‘you pinch (pl.)’	Quitepec Chinantex	Laver 1994:241

In interpreting these transcriptions we must ask ourselves what do the different authors mean to represent with the syllabicity mark. It quickly becomes apparent that its interpretation varies from author to author, some making no distinction between əC and C̩ or even simply a C surrounded by less sonorant consonants, either at the phonological level or phonetic level (Houlihan 1973). We are still left asking a number of questions: Are syllabic consonants underlyingly (phonologically) vowel-consonant sequences? Could it instead be the case that (some) syllabic consonants attach directly and uniquely to the syllable nucleus? Do syllabic consonants actually have vowels of some sort associated with them on the surface (phonetically)? If so, are these vowels always present? These questions are important as they relate directly to discussions concerning what can and cannot occur in a nucleus, onset or coda, and thereby to the typology of the syllable. The answers to these questions are also of crucial importance to phonological theories where there are very strict conditions on syllable structure and the presence versus absence of epenthetic vowels. Such a phonological theory will be discussed in §6.

2.1 Syllabic consonants in British English

In British English it is widely recognised that different styles produce different phenomena with respect to syllabic consonants (Roach, Sergeant and Miller 1992). In semi-formal registers /n/, /l/ and /r/ may be syllabic, the latter only in some rhotic dialects. Such syllabics are found in post tonic stress positions and almost always[1] before a morpheme boundary, for example [bɒtl̩] 'bottle' and [bɒtl̩iŋ] 'bottling' (Roach 1991:79). Syllabic consonants in semi-formal registers are often described as obligatory, in as much as it is deemed to be a mispronunciation to say, for example, [pi:pəl] rather than [pi:pl̩] for 'people' (Jones 1976:56).

[1] It is indeed clear that the vast majority of syllabic consonants in this register are found immediately before a morpheme boundary (either word internally e.g. ‚buttoning’, or word finally e.g. ‚button’). However, there are a small number of cases where a potential syllabic consonant may appear word internally and not at a morpheme boundary. The first set of such cases consists of words such as 'present', 'Arnold', where according to Wells (2000) the /n/ in the former and the /l/ in the latter are typically pronounced as syllabic. A pilot study for this paper found that the final stops (i.e. following the potentially syllabic consonants) in these words were virtually always dropped in the register under investigation (potentially a result of the carrier sentence where target words were followed by consonant initial words cf. Cruttenden 1994:261). The second set of cases consists of words ending in [iː ('-y') such as 'history', 'botany', 'chicory' and 'litany'. In these cases it is arguable that the '-y' is a separate morpheme on the grounds of alternations such as 'botany' ~ 'botanic' and thus the potential syllabic consonant is indeed adjacent to a morpheme boundary. However, this still does not help with 'chicory'. A third, disparate set of words is suggested by Rubach (1977:29) and includes 'cabinet', 'definite', 'obsolete'. Even if we accept Rubach's claim that these words contain syllabic consonants (he provides no phonetic evidence, and as we shall see below, syllabic [n̩] is rarely found after non-coronal consonants), it may be possible to claim that, although the endings '-et', 'ite', '-ete' and '-ute' may no longer be productive derivational morphemes, historically there is a morpheme boundary immediately following the potential syllabic consonant.

In fast and/or casual registers, /m/, /ŋ/ and even obstruents have been claimed to be syllabic, arising through processes of assimilation and elision. Such syllabics can be found in pre stress positions and domain initially, and their syllabicity is entirely optional. Examples include [s̩pəʊz] (Roach, Sergeant and Miller 1992: 476), [prob̩bleɪ] (Sivertsen 1960, quoted in Bell 1978:185), [k̩juː] (Jones 1959, quoted in Roach, Sergeant and Miller 1992: 476), [f̩ˈgɒʔ] (forgot) and [n̩ʔ ˈmaʔtʃ] (not much) (Beaken 1971, quoted in Wells 1982:321). In neither semi-formal, nor fast/casual registers may syllabic consonants be themselves stressed[2].

In this paper I examine only syllabic /n/ and /l/ in semi-formal registers of non rhotic speakers of British English. In the following section I describe a series of experiments designed to investigate whether segments which are transcribed as syllabic consonants ever have vowels (schwas) associated with them, and how syllabic consonants differ from their onset and coda counterparts in terms of duration and formant characteristics.

3 Experiment design

3.1 Subjects

In total the speech of eight subjects was investigated. All subjects were female, native English speakers with a university education reporting normal speech and hearing. All were naïve to the purpose of the experiment.

(2)	Subject	Age	Linguistic background
	1 (EW)	26	Has always lived in South London
	2 (SH)	34	First 18 years Birmingham area
	3 (EB)	26	First 18 years in Cornwall
	4 (RA)	50	First 18 years in Wiltshire
	5 (RF)	24	First 18 years in Co. Durham
	6 (ES)	22	Has always lived in North London
	7 (CL)	26	First 18 years Twickenham
	8 (JB)	31	0-9 London, 9-18 Hertfordshire

All subjects other than Nos. 2 and 5 spoke with a modified Southern British English accent showing close systemic similarities to RP English. Subject 2 spoke primarily with a Southern British English accent, although some vowels were monophthised in a manner atypical of Southern British English. Subject 5 spoke with a distinct North Eastern accent[3]. Initially attempts had been made to include only speakers of accents very similar to RP. This however proved impossible and so the only restriction placed on speakers was that they did not vocalize their /l/s as is typical of some English dialects ([bɒʔɰ] for 'bottle').

[2] Whilst this generalization holds true for most dialects and registers, we may note that some analyses of General American /r/ argue that /r/ can be both syllabic and stressed in words such as 'bird' [br̩d], curt [kr̩t] (Bloomfieldian transcription). We also note the description given by Holder (1972) in Wells (1982:573) of some Guyanese English words which have final stress, including 'bottle' and 'table'.

[3] The results of this speaker at no time differed significantly from the results of other speakers. Thus, although her dialect was not that under main consideration, her results were nevertheless included in the final analysis. That her results did not differ significantly from those results for speakers with RP-type accents suggest that the results of these experiments may be extendable to dialects other than RP, although such a claim would clearly require considerable further research.

Due to circumstances beyond my control not all subjects were used for all parts of the experiment. The results concerning the distribution of syllabic /l/ and syllabic /n/ following singletons were drawn from subjects 1-6, as were the results concerning duration and quality of the syllabic consonants versus their non syllabic counterparts. The results for the distribution of syllabic /l/ and syllabic /n/ following clusters were drawn from subjects 2, 3, 7, and 8.

3.2 Materials

A list of disyllabic words potentially containing syllabic /l/ or /n/ was drawn up from Wells' (2000) *Longman Pronunciation Dictionary*, Rockey's (1973) *Phonetic Lexicon* and an online pronunciation dictionary, *Beep*[4]. One set of words contained potential syllabic consonants preceded by consonant clusters, whilst a second set of words contained potential syllabic consonants preceded only by singletons. In order to control for contextual effects, potential syllabic consonants were preceded only by /p/, /t/, or /k/ as singletons. Thus words like 'suckle' and 'button' (potential syllabic preceded by /p/, /t/, or /k/ singleton) were included whilst words like 'hovel' and 'oven' (potential syllabic preceded by singleton other than /p/, /t/, /k/) were not. The consonant clusters in target words were restricted to (homorganic) nasal + (voiced or voiceless) stop clusters. Thus words like 'ankle', 'angle', 'Hendon' and 'Kenton' (potential syllabic preceded by nasal+stop cluster) were included whilst words like 'coastal' and 'silken' (potential syllabic preceded by cluster other than nasal+stop cluster) were not. A list of monosyllabic words containing /l/ and /n/ in domain initial (onset) and domain final (coda) positions was also drawn up from the same sources. This included words like 'moon' and 'leaf', but not words like 'blue', 'end' or 'elf' (where /l/ and /n/ form part of clusters)[5]. Word frequency was not controlled for (to do so would have resulted in a very restricted set of usable words).

Each target word was place within a carrier sentence ('How does _ translate?') to ensure that all target words occurred in the same prosodic context (i.e. within the word carrying main sentential stress) thus providing some control of durational effects due to syntax, semantics, emphasis and emotion (O'Shaughnessy 1981). The two lists were the combined and the sentence order then quasi-randomised such that sentences containing identical target sequences did not occur adjacent to one another.

3.3 Procedure

Each subject was asked to read the list of sentences three times at 'normal conversational tempo'[6]. Subjects were recorded onto minidisk in a sound proofed room using a Brüel and Kjær condenser microphone (Type 4165) and measuring amplifier (Type 2609). These recordings were digitised at 22,500 Hz using Goldwave and then imported in PRAAT where spectrograms for each sentence were produced and labelled. Labelling was done by visual inspection of the waveform and spectrogram, and by listening to the speech signal.

[4] ftp://svr-ftp.eng.cam.ac.uk/pub/comp.speech/dictionaries/beep.tar.gz

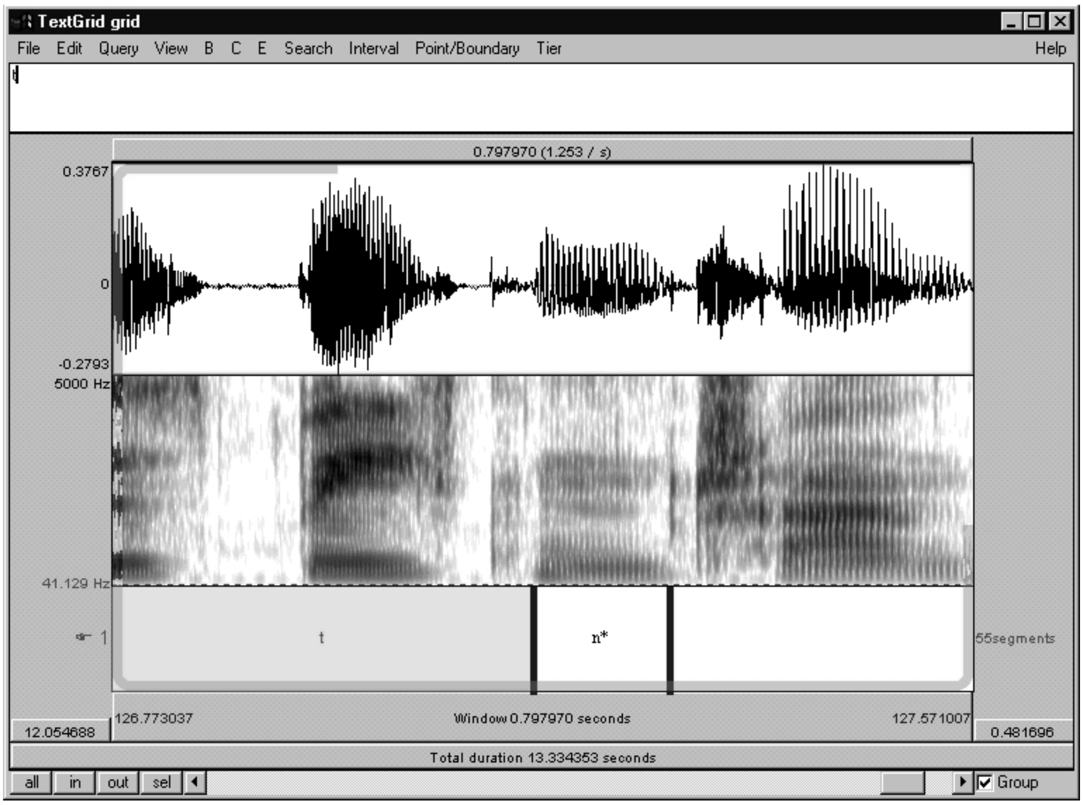
[5] Words ending in /l/ preceded by a high front vowel (e.g. 'feel' or 'pail') were also excluded from this experiment because speakers often lack clear intuitions as to whether such words are monosyllabic or disyllabic. In the latter case, such words could be argued to contain syllabic /l/s. Whilst this is a potentially interesting area of further research it will not be further addressed in this paper.

[6] No specific measures were taken to control for inter-speaker tempo variation

The segmentation guidelines were as follows: Fricative-sonorant boundaries (e.g. [haʊ də**z l**iːf…]) were established at the offset of frication noise and the onset of higher amplitude components in the periodic part of the signal. Vowel-sonorant (e.g. [**be**ɪkən]) boundaries were established by reference to discontinuities in the amplitude of the higher frequency components of the periodic signal, formant transitions and the presence of spectral zeros. Stop-sonorant and stop-vowel boundaries (e.g. [bɒ**tl̩**] or [beɪ**kə**n]) were established with reference to onset and offset of voicing, and the presence of higher amplitude components in the periodic part of the signal. Thus aspiration after the stop release was not included in the vowel (or sonorant) duration (contra Umeda 1976).

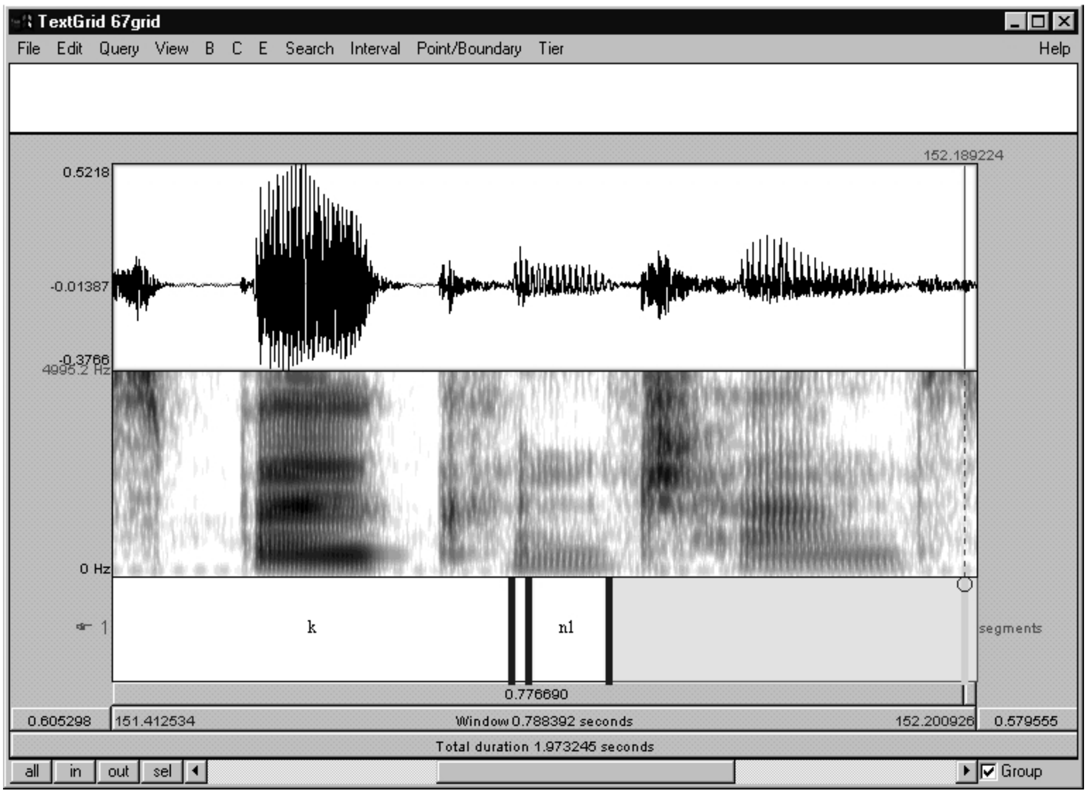
As can be seen in (3), a target consonant was labelled as syllabic if there was no evidence, either in the waveform or in the spectrogram, of a vowel of any sort. If there were visual cues to the presence of a vowel, as in (4), the target consonant was labelled 'post schwa'. In total 1705 sentences were recorded and labelled (some tokens were discarded due to noise interference on the recording e.g. speaker coughing or page turning). Duration and formant measurements were extracted[7] and all results were analysed using SPSS, where the threshold of significance was set at 95%.

(3) An example of labelling a syllabic consonant: 'beaten' [biːtn̩][8]



[7] Data was extracted using a PRAAT script. A random sample of these results were checked by hand and found to be accurate.
[8] 'C*' was my label for a syllabic consonant. Thus n* = [n̩]. 'C1' was my label for a post schwa consonant. Thus n1 = [n]. The onset of the preceding consonant is not marked.

(4) An example of labelling a non syllabic consonant: 'spoken' [spəʊkən]

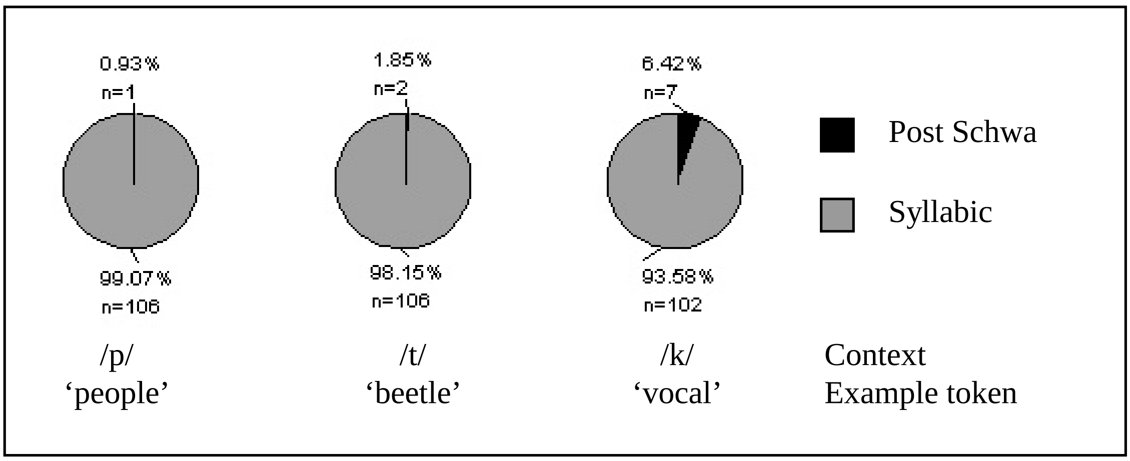


4 Experimental results

4.1 Results concerning distributional data

With regard to the distribution of syllabic /l/ and syllabic /n/ post singletons, both visual inspection and the results of statistical tests[9] confirmed that /l/ was found to be syllabic irrespective of context, whilst the distribution of syllabic /n/ was context dependent, being found primarily only after /t/, and not after /p/ or /k/.

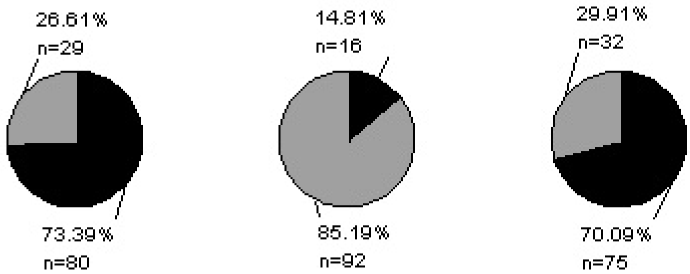
(5) Distribution of syllabic /l/ after /p/, /t/ and /k/, based on data from subjects 1-6



[9] For discussion of the tests used see Woods, Fletcher and Hughes (1986) and Rietveld and van Hout (1993).

The figures in (5) suggest that the distribution of syllabic /l/ is not random and a binomial statistic confirms that all distributions of syllabic /l/ deviate significantly from chance ($p<0.001$). Statistically there appear to be just significant differences between the three contexts, /p/, /t/ and /k/ (χ^2 (2) = 6.261, p = 0.044). However, this chi-squared was obtained from a table where 50% of the cells have an expected count less than 5. As a consequence the value obtained for p may be inaccurately low and therefore what appears to be just significant is actually unlikely to be so.

(6) Distribution of syllabic /n/ after /p/, /t/ and /k/, based on data from subjects 1-6



/p/	/t/	/k/	Context
'deepen'	'beaten'	'beacon'	Example token

The figures in (6) suggest that the distribution of syllabic /n/ is not random, and a binomial statistic confirms that all distributions of syllabic /n/ deviate significantly from chance ($p<0.001$). Statistically there are highly significant differences between the three contexts (χ^2 (2) = 93.914, $p<0.001$) and the pie chart patterns suggest that /n/ tends to be syllabic following /t/ and non-syllabic following /p/ and /k/.

These results show the first difference in behaviour between syllabic /l/ and syllabic /n/, by suggesting that place of context influences distribution of syllabic /n/ but not of syllabic /l/. /l/ appears to be always syllabic when it is preceded by any consonant and followed by a word boundary, whilst /n/ must be preceded by a coronal consonant to be syllabic[10,11].

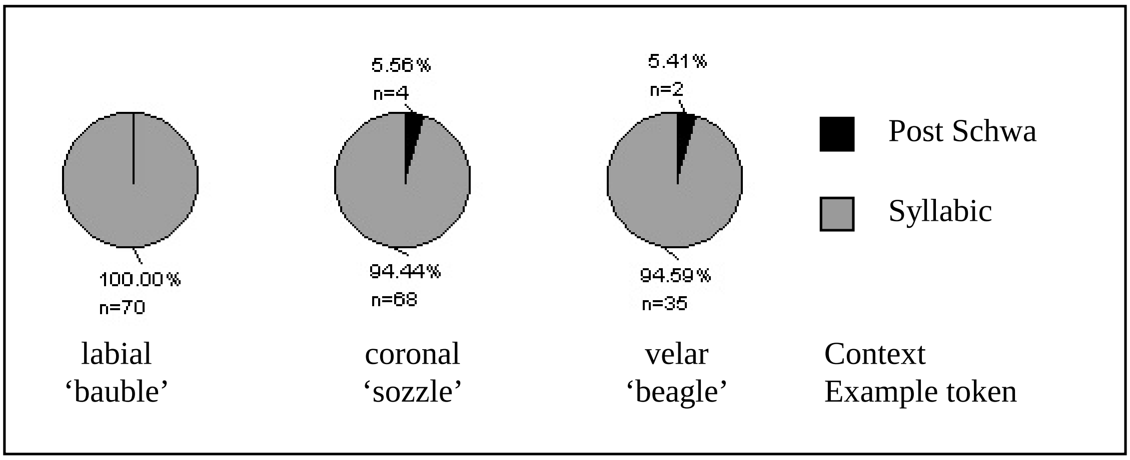
These results were supported by data from a pilot study (642 sentences) based on speaker 1 with an expanded set of contexts (/p, t, k, b, d, g, s, z, f, v/):

---

[10] It must be acknowledged that my results are statistical rather than categorical. Thus, for example, not all potential syllabic /n/s preceded by (homorganic) /t/ were actually realized as syllabic. One factor that may play a role in the realization of target consonants as syllabic is their relative frequency: it has been shown (Fidelholtz 1975) that relative frequency can significantly influence vowel reduction (the more frequent a word, the more likely it is to show vowel reduction). This being the case, it may be demonstrable that frequency also plays a role in total vowel absence, in other words, in the distribution of syllabic consonants. As reported in §3 word frequency was not controlled for in this set of experiments.

[11] As (6) shows, /n/ is syllabic approximately 25% of the time when preceded by /p/ or /k/. Readers might be tempted to speculate that these cases actually represent instances of assimilation i.e. [pm̩] instead of [pn̩], and [kŋ̍] instead of [kn̩]. However, this potentiality was borne in mind when labelling and care was taken to listen to the precise quality of the nasal and to label it accordingly. Thus the figures in (6) should be interpreted as representing cases of syllabic and non syllabic /n/ only: no cases of assimilation were found, which is perhaps not surprising given the relative formality of the recording environment and the standard descriptions that assimilation is more likely to occur in fast and casual speech than in formal speech (Roach, Sergeant and Miller 1992).

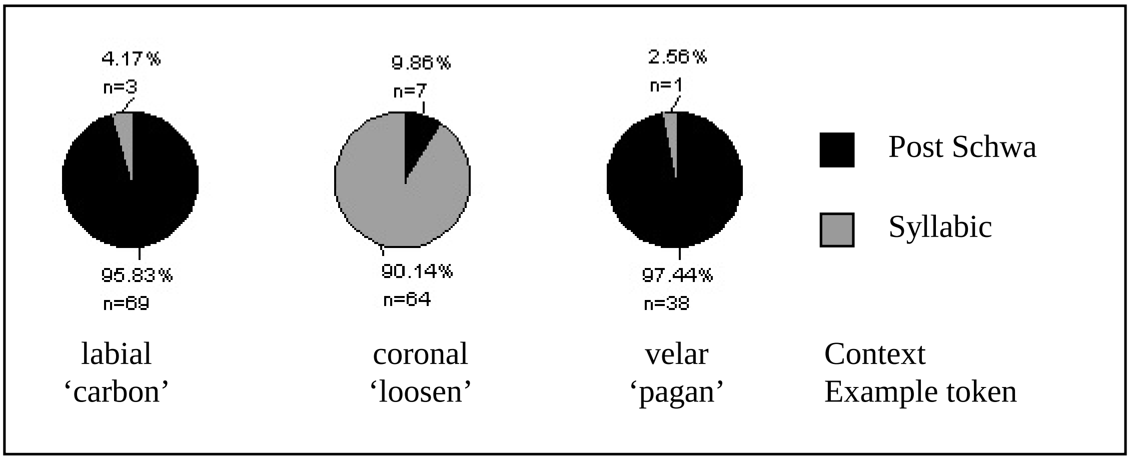
(7) Place: Distribution of syllabic /l/ after expanded set of contexts, classified by place of articulation (subject 1)



With this expanded set of contexts we see that there is *no* statistically significant effect of place on the occurrence of syllabic /l/ (χ^2 (2) = 3.99 p= 0.14). Again, this chi-squared was obtained from a table where 50% of the cells had an expected count less than 5, and thus it is likely that the value of p is inaccurately low.

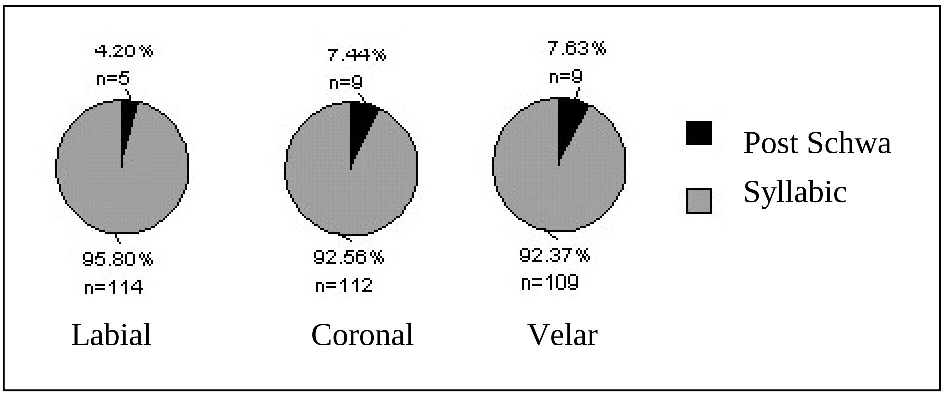
Turning to (8), below, again the data from this pilot study with an expanded set of contexts supports the conclusions reached on the basis of a restricted set of contexts. There is a highly statistically significant effect of place on the occurrence of syllabic /n/ (χ^2 (2) = 138.59, $p<0.001$) and the pie chart patterns show that it is primarily after homorganic i.e. coronal consonants that syllabic /n/ is found.

(8) Place: Distribution of syllabic /n/ after expanded set of contexts, classified by place of articulation (subject 1)



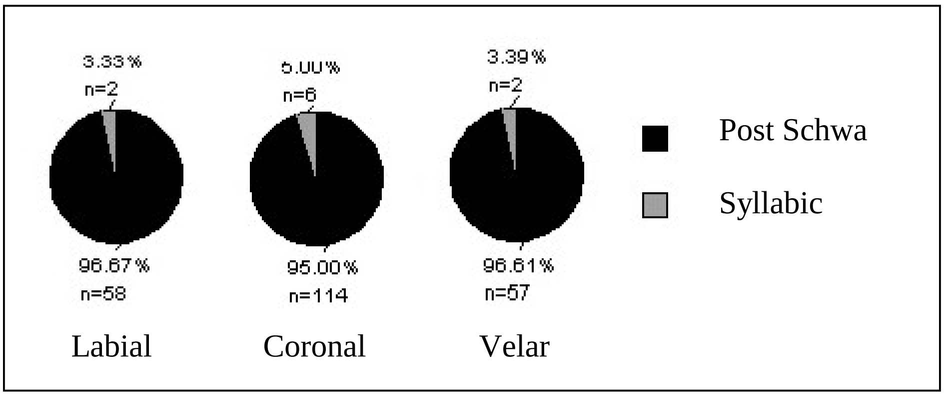
A second asymmetry in the behaviour of syllabic /l/ and syllabic /n/ is evident in their distribution following nasal+stop clusters. Both visual inspection and the results of statistical tests confirmed that /l/ was found to be syllabic irrespective of context, whilst /n/ was found never to be syllabic, not even when preceded by a /nt/, i.e. homorganic cluster.

(9) Distribution of syllabic /l/ after nasal+stop clusters (subjects 2, 3, 7, 8)



A binomial statistic shows that all distributions deviate significantly from chance ($p<0.001$). For target consonant /l/ there are not significant differences between the three contexts (χ^2 (2) = 1.469, p = 0.480).

(10) Distribution of syllabic /n/ after nasal+stop clusters



Again, for target consonant /n/, a binomial statistic shows that all distributions deviate significantly from chance ($p<0.001$) and that there are no significant differences between the three contexts (χ^2 (2) = 0.400, p = 0.819). This p may in fact be inaccurately low given that 33.3% of cells had expected counts of less than 5.

To conclude this section, if we consider and compare the results concerning the distribution of syllabic /l/ and syllabic /n/ following both singletons and clusters as presented in (5), (6), (9) and (10) we note two different patterns of behaviour. When a potential syllabic consonant is preceded by a singleton, context in the form of place of articulation, does affect the distribution of syllabic /n/: /n/ will be syllabic if the context is coronal. The distribution of syllabic /l/, however, is not influenced by place of context. When a potential syllabic consonant is preceded by a (nasal+stop) cluster we see a different pattern of behaviour: /n/ is never syllabic, not even when preceded by a coronal cluster, whilst /l/ continues to be syllabic irrespective of context.

4.2 Continuous data results: durations and formant characteristics

In this section I first present data regarding the mean duration and mean formant values for four different /l/ allophones: onset (word initial), coda (word final), syllabic and post schwa. We see that with regard to duration, coda /l/ is distinct from all other allophones, being significantly longer. With regard to formant values we see that onset /l/ is distinct from all other allophones, having a significantly lower F_1 and higher F_2.

In the second set of data we see that /n/ allophones do not pattern like /l/ allophones with respect to duration: onset /n/ is significantly shorter than all other allophones. Formant values for /n/ allophones were not investigated; there are no major differences in articulation for the different allophones and nasal formants are highly variable from one speaker to the next because of anatomical differences (Stevens 1998). Given this, any formant data gathered would be difficult to meaningfully analyse.

4.2.1 Results for /l/

(11) Summary of continuous data for /l/ based on subjects 1-6: Means (standard deviations)

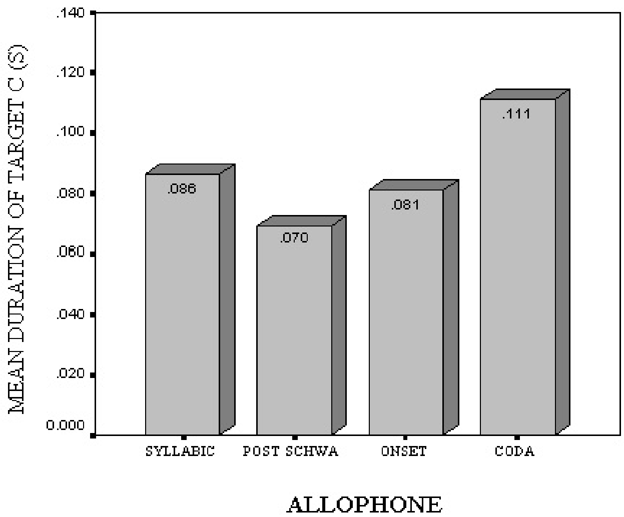
	Syllabic (N=314)	Post Schwa (N=10)	Onset (N=114)	Coda (N=116)
Duration (ms)	86 (18)	70 (17)	81 (17)	116 (27)
F_1 (Bark)[12]	4.62 (0.45)	4.52 (0.36)	3.90 (0.32)	4.41 (0.59)
F_2 (Bark)	9.03 (0.85)	8.82 (0.41)	11.13 (0.71)	8.76 (1.46)

NB. N = number of tokens

In (11) there appear to be differences in mean duration of the different /l/ allophones. A (one way) ANOVA test confirms that there is an effect of *target C allophone* on mean duration, ($F(3, 550) = 79.70$, $p<0.001$). A Tukey post hoc test shows that the duration of *coda* /l/ is significantly longer than syllabic /l/, post schwa /l/ and onset /l/ ($p<0.001$ in each case).

[12] All frequency measures are given in Bark (not Hz). The bark scale is perceptually more real than the Hz scale, and I have chosen to use it as I believe auditory perception to be at the heart of phonology (cf. Harris 2000).

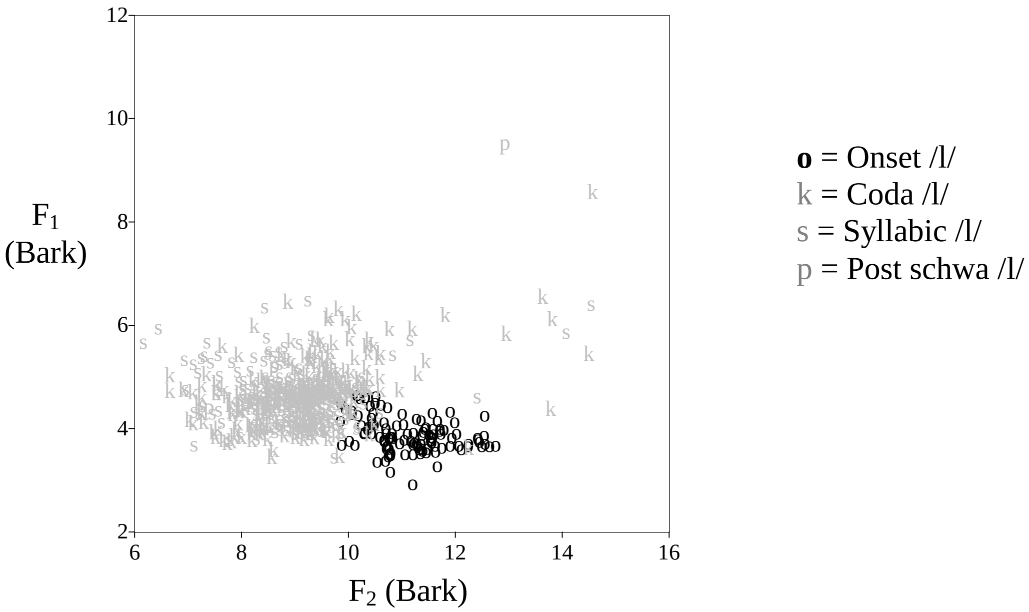
(12) Mean duration (in seconds) of /l/ allophones based on subjects 1-6



A Tukey post hoc test also shows that the duration of syllabic /l/ is significantly longer than post schwa /l/ (p=0.045). However, none of the other durational differences are significant.

If we now consider the formant characteristics of the four /l/ allophones, there appear to be differences in the means of both F_1 and F_2. An ANOVA test confirms that there is an effect of *target C allophone* on mean formant values ($F(3, 550) = 69.13$, $p<0.001$). Tukey post hoc tests show that the F_1 of onset /l/ is significantly lower, and the F_2 of onset /l/ is significantly higher than those of the other allophones ($p<0.001$ in each case). In other words, onset /l/ can be characterised as 'light' or 'clear', whilst coda /l/, syllabic /l/ and post schwa /l/ are 'dark', or velarized[13], as can be seen in (13).

(13) A plot of F_1 against F_2 for /l/ allophones based on subjects 1-6



[13] Quite what articulation gives rise to this acoustic quality is a matter of debate. See Cruttenden (1994) and Sproat and Fujimura (1993) for differing views.

4.2.2 Results for /n/

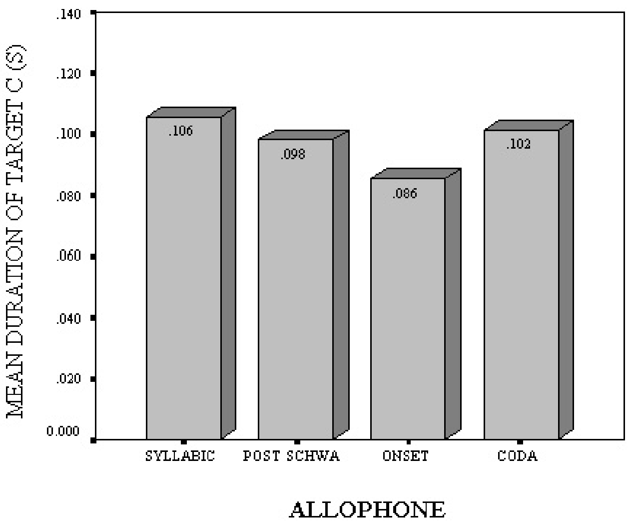
(14) Summary of durational data for /n/ based on subjects 1-6: Means (standard deviations)

	Syllabic (N=153)	Post Schwa (N=171)	Onset (N=116)	Coda (N=114)
Duration (ms)	106 (20)	98 (18)	86 (24)	102 (22)

NB. N = number of tokens

An ANOVA test shows that there is an effect of *target C allophone* on the mean duration of the four /n/ allophones (F (3, 550) 21.92, p<0.001) and a Tukey post hoc test shows that the duration of onset /n/ is significantly shorter than that of syllabic /n/, post schwa /n/ and coda /n/ (p<0.001 in all cases). Onset /n/ is on average 16 ms shorter than the other allophones of /n/ as can be seen in the graph below.

(15) Mean duration of /n/ allophones based on subjects 1-6



A Tukey post hoc test also shows that the duration of syllabic /n/ is significantly longer than the duration of post schwa /n/ (p = 0.006), though none of the other differences in duration are significant.

To summarise and conclude §4, what we have seen above is three ways in which syllabic /n/ behaves differently from syllabic /l/: (i) Distribution of syllabic /n/, but not syllabic /l/ is sensitive to the context's place of articulation; (ii) Distribution of syllabic /n/, but not syllabic /l/ is sensitive to the context's structural complexity i.e. whether or not it is preceded by a singleton or a cluster; (iii) The duration of syllabic /n/ is not significantly different from the duration of coda /n/ whilst it is significantly different from the duration of onset /n/. Syllabic /l/, on the other hand, is not significantly different from onset /l/ with respect to duration, yet it is significantly different from coda /l/. Thus syllabic /l/ displays a mirror image pattern with respect to syllabic /n/.

5 Discussion of phonetic data

In this section I present data published by earlier authors on the durational and formant characteristics of different allophones of English /l/ and /n/ by way of comparison with my data above. This will lead to a discussion of the phonological value of durational and formant cues and the relationship more generally between continuous phonetic data and categorical phonological structures.

The following data is taken from Barry (2000), Byrd (1993), Lehiste (1964), Sproat and Fujimura (1993) and Umeda (1976)[14]. Data presented above in (11) and (14) is also included here for convenience.

(16) A comparison of /l/ allophone formant values: F_1 and F_2[15]

	Author	Syllabic	Onset	Coda
F_1 (Bark)	Toft	4.62	3.90	4.41
	Lehiste	4.35	3.08	4.57
	Barry	-	-	-
F_2 (Bark)	Toft	9.03	11.13	8.76
	Lehiste	7.15	8.21	7.21
	Barry	8.00	9.92	8.33

(17) A comparison of /l/ allophone formant values: F_2 - F_1[16]

	Author	Syllabic	Onset	Coda
F_2 - F_1 (Hertz)	Toft	630	1146	606
	Lehiste	355	655	340
	Sproat & Fujimura	-	1077	657

[14] The experiments from which data has been drawn varied in their scope and aims and thus are not directly comparable with my own data: data collection conditions (e.g. stimuli, speaker sex and number) vary from paper to paper. Barry (2000) is based on acoustic and articulatory (EPG) data from 3 speakers of RP English for formant data and 2 speakers for duration data. Their sex is not specified. Byrd (1993) is based on acoustic data drawn from the TIMIT corpus (630 speakers of American English, with 974 tokens of syllabic /n/ and 9660 tokens of non-syllabic /n/). Lehiste (1964) is based on acoustic data from 5 speakers of Midwestern American English (sex unspecified). Sproat and Fujimura (1993) is based on data from 5 speakers, 2 of which were female. One male informant is described as a 'speaker of British English with a fair amount of American English influence' whilst all other subjects were speakers of Midwestern American English. Umeda (1976) is based on data from 1 male speaker of American English. Only Lehiste and I use a regular sentence frame format; data from other authors collapses different prosodic and sentence positions for target segments.

[15] Lehiste and Barry actually quote their formant values in Hz. I have converted their values into Bark using the formula $7 \ln (x/650 + \sqrt{(1 + (x/650)^2)})$ as used in the speech software PRAAT. The data quoted here from Barry specifically excludes prevocalic syllabic /l/ in order to ease comparison with my own data.

[16] This table is given in Hz as Sproat and Fujimura only provide F_2-F_1 values in Hz: to do so in Bark would be meaningless as the Bark scale is a logarithmic scale. The formula 650 sinh (x / 7), as given in PRAAT, was used to convert Bark values into Hz values.

The data in (16) and (17) concurs with descriptions given by Spencer (1995:214-6), Roach (1991:79) and Gimson (1989:202) who assert that syllabic /l/, like coda /l/ is always dark. However, before we adopt this claim wholesale we must consider the fact that, at least with respect to my own data, the dark quality of syllabic /l/ may be a construct of my experiment design[17]: due to the nature of the carrier phrase, syllabic /l/ was always followed by a consonant ('How does _ translate?'), and never by a vowel. Barry (2000) provides some evidence that, at least in RP English, syllabic /l/ is regularly 'light' and onset-like in its quality when followed by a vowel initial word, rather than a pause or a consonant initial word. Here it is worth noting that syllabic /l/ in Lehiste's data was followed by a vowel initial word (her carrier phrase was 'Say the word _ instead'), and still syllabic /l/ remained 'dark'. This may turn out to be a dialectal difference between the two Englishes (as, for example, in the case of intervocalic /l/ which is clear in RP and dark in GenAm, according to Wells (1982:74)).

I believe that further investigation into the quality of a syllabic /l/ is worthwhile. Quality differences such as 'dark' and 'light' may provide valuable cues to the parsing of phonological strings. Such quality differences may demarcate prosodic boundaries and therefore these cues should have a place in the phonological representation. Given these assumptions it is vital to establish the quality of syllabic /l/ and whether or not this varies with context.

Let us now turn to the issue of duration and its place in phonological representation. Clark and Yallop (1995:67), Jones (1959:136) and Price (1980) have claimed that duration is a cue to syllabicity; syllabic consonants are purported to be longer than their non syllabic counterparts. This claim, however, is not immediately substantiated by the data presented in (18) and (19)

(18) A comparison of /l/ allophone durations[18]

	Author	Syllabic	Onset	Coda
Duration (ms)	a. Toft	86	81	116
	b. Lehiste	253	-	-
	c. Barry	81	95	78
	d. Umeda	-	66	-

(19) A comparison of /n/ allophone durations[19]

	Author	Syllabic	Onset	Coda
Duration (ms)	a. Toft	106	86	102
	b. Umeda	-	71	(i) 81 (ii) 48
	c. Byrd	80	59	59

First, it is clear that duration of onset vs. coda sonorants i.e. the non-syllabic sonorants is not identical, with the exception of the data from Byrd in (19c). The process of domain final lengthening of consonants is well established in the literature (e.g. Crystal and House 1987a,

[17] This was pointed out to me by Martin Barry (p.c.).

[18] This figure differs from the average duration quoted by Lehiste (actually 247 ms, p.31) in that it is calculated on the basis of the duration of syllabic /l/ when preceded by /p/, /t/ and /k/ only; Lehiste's average is based on syllabic /l/ preceded by 18 different obstruents but I have extracted the /p/, /t/ and /k/ values to assist comparison with my data.

[19] Two values for coda duration are quoted by Umeda. (i) = duration of coda /n/ when followed by a pause, whilst (ii) = duration of coda /n/ when followed by a vowel initial word.

Cutler et al 1997) and it may be that this can account for the longer coda durations in (18a), (19a) and (19bi). Whether syllabic consonants should be compared to their onset or coda counterparts is unclear. Given that syllabic consonants are inevitably domain final, perhaps codas do provide the best grounds for comparison. However, the results from (18a) and (19c) suggest that a direct comparison is not possible. Crystal and House (1987b) show that all consonants are lengthened under stress and it could be argued that this accounts for the data in (18a): my stimulus words for coda consonants were monosyllabic, whilst for syllabic consonants they were disyllabic and thus my coda consonants were elicited as part of stressed monosyllables while my syllabic consonants were not in stressed syllables (syllabic consonants are necessarily unstressed in English). In contrast, coda /n/s in Byrd (1993) are not restricted to monosyllabic stressed words (though no detail is provided as to how many tokens were indeed unstressed).

So far all that is clear is that the data does not yet support the strong claim that all syllabic consonants are necessarily cued by longer durations. It may be that claims concerning the longer duration of syllabic consonants vis a vis their non syllabic counterparts are based on a comparison of syllabic consonants vs. non syllabic consonants *in a cluster*, though this is nowhere stated explicitly. Thus the two /l/s in e.g. [bʌkl̩] vs. [kli:n] should be compared with each other rather than the two /l/s in [bʌkl̩] vs. [li:n]. The collection of such data for English awaits further research, however, it is perhaps interesting to consider some already available data from Icelandic.

Icelandic, much like English, has word final obstruent-sonorant sequences. However, whilst such word final sonorants are described as syllabic in English, they are not described as such in Icelandic (Ito 1987, Árnason 1980).

(20) Some word final obstruent sonorant sequences in English and Icelandic

English	Icelandic
/æpl/ apple	/apl/ strength
/bætl/ battle	/fatl/ downfall
/hækl/ hackle	/fʏkl/ bird
/bʌtn/ button	/ɛɪtn/ one

If longer duration is a cue to syllabicity of consonants (Clark and Yallop, and Jones do not specifically restrict their claims to English) then the apparently non syllabic final sonorants of Icelandic should not be especially longer than their non syllabic word internal consonants. In other words, the /l/ in /apl/ 'strength' should be no longer than the /l/ in /plau:r/ 'blue', providing we can control for processes like phrase final lengthening. The following data is taken from Garnes (1976):

(21) Sonorant duration in Icelandic

Position	/l/	/n/
Singleton onset	[l]ög 98 ms	[n]einn 76ms
Part of word internal cluster of increasing sonority[20]	ep[l]i 64 ms	vak[n]a 69 ms
Word final, post obstruent	gut[l] 189 ms	vat[n] 260 ms

[20] pl- and kn- are both possible word initial onsets in Icelandic. However Garnes does not provide data for words with such onsets.

Whilst this data from Garnes must be treated with caution (her corpus contains no other words which either end in /tn/ or /tl/ or begin with a singleton /l/ or /n/), it would appear that word final post obstruent sonorants in Icelandic are considerably longer than their onset counterparts, either as singletons or as part of a cluster. One possible conclusion might be that Icelandic does in fact have syllabic consonants. Thus the claim that syllabicity is cued by duration could be maintained. Alternatively we could abandon the assumption that there is a direct relationship between longer duration and syllabicity. Clearly this is an area for further research.

For now, however, it is to an explanation of the behaviour observed and described in §4 and §5 that we turn. In the following section I outline a phonological framework where there are severe restrictions on the syllable template and strict conditions govern the occurrence of epenthetic schwas. Using this framework, in §7 I then turn to an analysis of the data presented above.

6 A phonological framework

Government Phonology (GP) aims to provide a non-arbitrary account of phonological events by replacing the rule component of a phonology with a finite set of universal principles and parameters (contra Bromberger and Halle 1989). Paralleling Chomsky's 'principles and parameters' approach to the syntax of natural languages (Chomsky 1981, 1982, Culicover 1997), the different phonological systems of languages are captured through different combinations of parameter setting.

Like any school of thought Government Phonology also comes in a variety of flavours (for example, Harris 1994, Scheer 1998, Cyran 2000). With regard to constituency I shall primarily be following work by Kaye, Lowenstamm and Vergnaud (1990), Charette (1991a) and Harris (1994) whilst with regard to melody my work shall be based on Harris and Lindsey (1995).

6.1 Prosodic structure

The only constituents available in Government Phonology are the Onset, the Nucleus and the Rhyme, each of which may be maximally binary branching (Kaye, Lowenstamm and Vergnaud, henceforth KLV, 1990:198-199). No 'syllable' constituent is recognised (KLV 1990:200-201, Harris 1994: 45-46), but rather, domains are constructed of iterated Onset-Nucleus pairs: neither a Nucleus, nor an Onset may form a domain on its own, nor may two Onsets (or two Nuclei) follow one another, without an intervening Nucleus (or Onset) (Harris 1994:160). No 'Coda' constituent is recognised either (Kaye 1990a): Any word final consonant is instead attached to an Onset, which is necessarily followed by a (licensed, empty) nucleus (see below).

Constituents are attached to the melody (i.e. segmental structure) via a timing unit, traditionally referred to as a skeletal point (Kaye 1989:125-139, Harris 1994:33-41). Thus phonological quantity, as manifested for example in the difference between light and heavy diphthongs, geminate and non-geminate consonants, and of course long and short vowels, is captured independently from both the prosodic and melodic tiers. To the best of my knowledge no work has been done within GP on establishing a relationship between the phonological and phonetic value(s) of skeletal points, in contrast to the now quite substantial literature available on the phonetic interpretation of the mora (e.g. Beckman 1982, Hubbard 1995, Barnes 2001). Thus one aim of this paper is to provide data for investigating this relationship between phonology and phonetics, specifically within a GP framework. Of course, whilst attempts at establishing a relationship between phonological structures and phonetic durations are, I believe, worthwhile, it remains clear that we must maintain a relational sense of time, and not an absolute one.

6.2 Melody

Language is primarily and auditory system of symbols. In so far as it is articulated it is also a motor system, but the motor aspect is clearly secondary to the auditory. In normal individuals the impulse to speech first takes effect in the sphere of auditory imagery and is then transmitted to the motor nerves that control the organs of speech. The motor processes and the accompanying motor feelings are not, however, the end, the final resting point. They are merely a means and a control leading to auditory perception in both speaker and hearer…Hence, the cycle of speech… begins and ends in the realm of sounds.

(Sapir 1921:18-18, quoted in Harris and Lindsey 2000:185)

Whilst the claim that linguistic units of sounds are decomposable into smaller units was adopted the best part of a century ago, the nature of these subparts remains a topic of discussion. In those approaches which adopt traditional features, such as [+/- front], [+/- anterior] as the smallest phonological units, articulation is elevated to unwarranted levels of importance given the wealth of evidence showing that speech production is parasitic on speech perception[21]. The primacy of articulation is undermined not least by the fact (all too often dismissed) that the same acoustic signature can be achieved by very different articulatory means.

In contrast to many feature based approaches, Government Phonology takes seriously the relationship between phonological information and information in the speech signal. Sounds are composed of one or more monovalent elements, which although not in themselves acoustic events, are directly mappable onto gestalt patterns in the acoustic signal.

[21] Consider, for examples, studies showing how acquired deafness and distorted auditory feedback can impair speech production e.g. Perkell et al. 2000.

(22) Elements in Government Phonology (Harris 1994:140, 1995:66)

Symbol	Acoustic Pattern and notes
A	Central spectral energy mass (convergence of F_1 and F_2). Present in uvulars and pharyngeals.
I	Low F_1 coupled with high spectral peak (convergence of F_2 and F_3). Present in palatals and palatalized consonants.
U	Low spectral peak (convergence of F_1 and F_2). Present in labial vowels and consonants.
@	Dispersed formant structure i.e. no salient spectral peak (a baseline on which other resonances are superimposed). Present in velar consonants
R	Set of formant transitions associated with coronals[22]
h	Aperiodic energy, the result of a narrowed stricture which produces turbulent airflow.
?	Abrupt and sustained decrease in overall amplitude, independently manifested as a glottal stop
N	Low frequency broad band murmur
H	High fundamental frequency, stiff vocal chords
L	Low fundamental frequency, slack vocal chords

Whilst this maximal element inventory is considerably smaller than any standard feature inventory (10 vs. 22 subsegmental units[23]), thereby substantially constraining the number of predicted combinations (1023 vs. 4, 194, 304)[24], work continues within the GP tradition to reduce the element set further (Jensen 1994, Ploch 1999)[25]. However, for the purposes of this paper I shall continue to use the 10 element set above, not least because any reduction in the set of elements risks also reducing the potential for directly mapping elements onto acoustic-auditory cues. I propose the following representations for the consonantal inventory of English:

[22] In fact, a single signature pattern for the element R has proved elusive, although see Kang's 1999 (non GP) work on the acoustic properties of coronals. An additional problem arises in proposing that coronality has an independent element on a par with others: the special properties of coronals (for example its propensity to assimilate) fail to be captured (Harris 1995:67-68).

[23] This total (22) is taken from Halle and Clements (1983).

[24] (2^{10}) -1 = 1023 (empty set is excluded from total), 2^{22}= 4,194,304 (no empty set, with all features fully specified).

[25] This work reminds us to ask what is it that we are trying to capture with our feature/element sets. If we are trying to capture and characterise all and only phonemic contrasts, we may be aiming to reduce the set of features/elements such that only about 100-150 combinations can be generated, given that the largest known inventories contain approximately this number of segments e.g. !Xũ, with 141 segments (Crystal 1987;165). If this is our aim, a 10 element system results in considerable overgeneration. If, however, we wish to characterise all *cues* used in expressing these contrasts a larger set of combinations will be necessary. Quite how large a set of cues is necessary is a matter for empirical research: a universal set of acoustic-auditory cues is far from general acceptance (not least because of the historical bias towards articulation in phonetics and phonology). In reference to English Harris (1994) gives 10 cues, but more may be necessary to capture salient signal cues in other languages.

(23) Internal representation of the consonantal inventory of Southern British English (broadly based on Harris 1994)

Bilabial	Labio-dental	Dental	Alveolar	Post-Alveolar	Palatal	Velar	Glottal
p (@.U.ʔ.h.H) b (@.U.ʔ.h)			t (@.R.ʔ.h.H) d (@.R.ʔ.h)			k (@.ʔ.h.H) g (@.ʔ.h)	
				tʃ (@.I.R.ʔ.H) dʒ (@.I.R.ʔ)			
m (@.U.ʔ.N)			n (@.R.ʔ.N)			ŋ (@.ʔ.N)	
	f (@.U.h.H) v @.U.h)	θ (@.R.h.H) ð (@.R.h)	s (@.R.h.H) z (@.R.h)	ʃ (@.I.R.h.H) ʒ (@.I.R.h)			h (@.h)
			ɹ (@.R)		j (@.I)	w (@.U)	
			l (@.R.ʔ)				

As can be seen from the table above, PEs often contain more than one element. However, there are certain restrictions on elemental combination (also known as fusion, and indicated by the symbol '.'). Universally each element may maximally occur once in any given expression. Each expression may also contain maximally one Head element: in principle any element may take the role either of Head (by convention the Head of a PE is underlined) or of Operator. Headedness represents the preponderance in the acoustic signal of the given element's signature over the signature of other elements[26]. The elemental composition of language specific sound inventories are established by examining the acoustic signal and matching the patterns observed with the patterns which represent different elements as described in (22) above. Thus, elemental composition is not stipulated but, at least in principle, established on independent grounds (Williams 1997).

The element @ has special properties which warrant some discussion. It does not display the sort of active resonance properties that other resonance elements (A, I, U, R). Furthermore it is argued to be present in all expressions, although its auditory-acoustic signature is only manifested when @ is head of the given expression (Harris 1994:108-113). A PE that contains only @, as an operator, is known as an empty expression, and the nucleus to which such an expression is attached is known as an empty nucleus. Special principles govern the manifestation of such empty nuclei and these are discussed below.

[26] This is in fact an idealised scenario, as empirical support for this claim has so far proved elusive.

6.3 Empty nuclei

Many languages exhibit vowel-zero alternations. The vowels involved in such alternations are typically reduced, for example the alternation ɨ ~ ∅ in Moroccan Arabic (Kaye 1990b), and ə ~ ∅ in French (Charette 1991a). In Government Phonology, vowels which have this special property of alternating with zero are argued to be underlyingly empty (or in terms of the theory of elements outline above, to contain only operator @) and their interpretation is subject to the Empty Category Principle (ECP) and concomitant parameters:

(24) The Phonological Empty Category Principle

(Charette 1990:235, precise wording here taken from Kaye 1995:295)

A p-licensed (empty) category receives no phonetic interpretation

There are four potential circumstances under which an empty category, or specifically here, an empty nucleus, can be p-licensed.

(25) (Parametric) Conditions on p-licensing

(Kaye 1995:295, precise wording here my own)

A p-licensed (empty) category may receive no phonetic interpretation iff:
i. it is magically licensed
ii. it is within an Inter Onset governing domain
iii. it is domain final
iv. it is properly governed

Magic licensing (Kaye 1991) is concerned with nuclei preceding S+C clusters. As such it is not relevant to the discussion at hand and therefore no further consideration shall be given to it here. Inter Onset Government (Gussmann & Cyran 1998, Lee 1999) is a governing relationship which may, in languages where this particular parameter is ON, be contracted between two onsets separated by a nucleus if certain substantive constraints are met (see §6). The domain final parameter, referred to in (i) above, is ON in those languages which allow apparent word final consonants (e.g. English) and OFF in those languages where words may only end in vowels (e.g. Cayuvava, Hawaiian). Given that GP holds that phonological strings consist of Onset-Nucleus pairs, those words, which appear to end in consonants, actually end in a nucleus. This nucleus is empty (it has no melodic content) and because it is silent it must be (by stipulation) p-licensed. By convention, a p-licensed category is indicated by underlining.

(26) Domain final p-licensing: representation of 'dog'

```
O   N   O   N   ◄────── A domain final
|   |   |   |           p-licensed nucleus
x   x   x   x̲
|   |   |
d   ɒ   g
```

Finally, an empty category may remain without interpretation if the conditions for Proper Government are met.

(27) Proper Government

(Kaye 1995:295, precise wording here taken from Charette 1998:170)

A nucleus α properly governs and empty nucleus β iff:
i. α and β are adjacent on the nuclear projection
ii. α is not itself p-licensed
iii. α is not a government licensor (for its onset).

(28) Government Licensing (Charette 1990:242)

For a governing relation to hold between a non-nuclear head A and its complement B, A must be licensed to govern by its nucleus.

The clause in (27iii) refers to situations where a consonant cluster intervenes between the two nuclei. Let us first examine the simpler case of Proper Government when the empty nucleus is preceded by a singleton. An underlyingly empty nucleus is realised as zero when it is properly governed by a following unlicensed nucleus. If Proper Government fails to apply the empty nucleus is phonetically interpreted.

(29) Bulgarian singular ~ plural alternation (data taken from Cowan and Rakusan 1985)

bobər ~ bobri	'beaver'	lavər ~ lavri	'laurel'
kosəm ~ kosmi	'hair'	teatər ~ teatri	'theatre'

<u>Proper Government fails.</u>
N^2 and N^3 are adjacent on the nuclear projection, but N^3 is itself p-licensed. Consequently

	N^1		N^2		N^3
	|		|		|
O	N^1	O	N^2	O	N^3
|	|	|	|	|	|
x	x	x	x	x	<u>x</u>
|	|	|		|	
b	o	b	[ə]	r	

<u>Proper Government succeeds.</u>
N^2 and N^3 adjacent on the nuclear projection, N^3 is not itself p-licensed (we know this because it has melodic content), and N^3 is not preceded by a

	N^1		N^2		N^3
	|		|		|
O	N^1	O	N^2	O	N^3
|	|	|	|	|	|
x	x	x	<u>x</u>	x	x
|	|	|		|	|
b	o	b		r	i

Now let us consider the more complex case where a consonant cluster does intervene between the two nuclei.

(30) Failure of p-licensing when consonant cluster intervenes. Fr. 'secret'

```
        N¹              N²
        |               |
O¹      N¹      O²      N²
|       |       |       |
x       x       x   x   x
|               |   |   |
s      [ə]      k   r   ɛ
```

<u>Proper Government fails</u>
N^1 and N^2 are adjacent on the relevant projection
but an consonant cluster intervenes
O^2 requires a license to govern its complement and this it gains N^2 from, leaving N^2 unable to Properly Govern N^1

Finally we note a third situation: that when a empty nucleus which could be properly governed and thereby p-licensed, is preceded by a consonant cluster. In such instances, Proper Government and Government Licensing interact, though the nature of their interaction is parametrically set.

(31) Parametric interaction of Government Licensing and Proper Government

French: (word internally) only non p-licensed empty nuclei may act as Government Licensors

```
        N¹              N²          N³
        |               |           |
O¹      N¹      O²      N²    O³    N³
|       |       |       |     |     |
x       x       x   x   x     x     x
|       |       |   |         |     |
l       i       b   r  [ə]    m     ã
```

<librement> 'freely' [librəmã], *[librmã]

The conditions for Proper Government of N^2 by N^3 are met, but because of the parameter regarding the interaction of Proper Government and Government Licensing, N^2 is nevertheless phonetically realised.

Polish: (word internally) p-licensed empty nuclei may act as Government Licensors

```
        N¹              N²          N³          N⁴
        |               |           |           |
O¹      N¹      O²      N²    O²    N³    O²    N⁴
        |       |       |     |     |     |     |
        x       x   x   x     x     x     x     x
        |       |   |   ‾     |     |     |     ‾
        u       b   r         d     a     c
```

<ubrdac>, 'imagine'
[ubrdac], *[ubrədac]

The conditions for Proper Government of N^2 are met, and N^2 may indeed remain silent because of the different parameter setting in Polish vis a vis French.

What is interesting about the examples in (31) is the interaction of two principles of Government Phonology (for more details see Charette 1991b). We have spent some time on this interaction as we shall see a similar case of interaction, though between two different principles, when we come to the phonological analysis in §6 and §7.

Overall it may be said that the key issue to have taken from this section is that there are strict principles governing the (non) interpretation of empty nuclei. Unless certain conditions are met empty nuclei are phonetically interpreted as reduced vowels, typically as schwa.

7 Towards a phonological analysis

As we saw above, 'syllables' in GP are made up of Onset-Nucleus pairs. Thus the syllabification of a world like 'button' necessarily includes a nuclear constituent between the /t/ and the /n/[27]. Furthermore, we have seen that GP has strict principles governing the non interpretation of nuclei; only p-licensed empty nuclei can remain without phonetic interpretation. Given this let us consider the predictions we can make concerning the realization of a word like 'button'

(32) Phonological structure for 'button'

O	N^1	O	N^2	O	N^3
|	|	|	|	|	|
x	x	x	x	x	$\underline{x}$
|	|	|		|	
b	ʌ	t	?	n	

What can we predict about the (non) interpretation of N^2? Are any of the conditions on p-licensing met? If they are, we predict N^2 to be silent. If none of the conditions on p-licensing are met we predict N^2 to be phonetically interpreted.

N^2 is not domain final, so it is not p-licensed through parametrically set domain final p-licensing. N^2 cannot be Properly Governed: it is followed by a empty nucleus which is itself p-licensed i.e. N^3. N^2 cannot be Magically Licensed (we are not dealing with an S+C cluster). Our only hope for meeting one of the conditions on p-licensing now lies with Inter Onset Government.

(33) Inter Onset Government

An Inter Onset Governing relationship may be contracted between an onset α and onset β iff:
i) α and β are adjacent at the Onset Projection
ii) α is a good Inter Onset governor and β a good Inter Onset governee, in terms of substantive constraints.

Note 1: An Inter Onset governing relation may be head final (e.g. Korean, Heo 1994) or head initial (e.g. Polish, Gussmann & Cyran 1998). The direction of Inter Onset Government is parametrically set.
Note 2: The substantive constraints on Inter Onset Government are similar to those for Constituent Government[28]

[27] Of course we are assuming that /tn/ cannot form a branching onset on the grounds, for example, that no English word could begin /tn/. For a principled GP explanation of why this is not possible see KLV 1985.
[28] The precise nature of substantive constraints for Inter Onset Government is not yet entirely clear. They may turn out to be linked to the direction of Inter Onset Government, head-final I-O government implicating Transconstituent-like substantive constraints, and head initial I-O government implicating Constituent-like substantive constraints (Toft 1999) but this is a subject for further research. For the purposes of this paper we

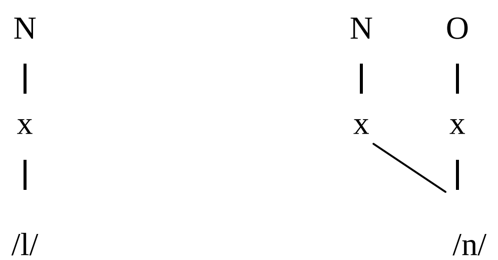
The substantive constraints on Constituent Government stipulate that the Governor be headed and more complex than the Governee, and that the Governor and Governee share at most one element other than the element @ (Kaye, Lowenstamm & Vergnaud 1990, Harris 1990). Given these constraints it is not possible for the /t/ and the /n/ in 'button' to enter into an Inter Onset governing relationship: the elemental representation of /t/ is (@.h.ʔ. R. H)[29], whilst for /n/ it is (@.R.ʔ.N). The two expressions have (@), (R) and (ʔ) in common i.e. too many elements in common, and thus an Inter Onset Governing relationship is not entered into. For similar reasons Inter Onset Government is ruled out between the stop and the liquid in a word like 'bottle'.

Here it is also worth noting that the acoustic result of Inter Onset Government (IOG) sounds very much like a branching onset30, and not like one consonant followed by a syllabic consonant. For example, one case in English where it has been suggested that Inter Onset Government may be at work is in the word 'chocolate', which is typically pronounce [tʃɒklət] but may be pronounced [tʃɒkələt] (Wells 2000:137). It is proposed that IOG is active when the former pronunciation is chosen, and inactive when the latter is opted for, thus resulting in an epenthetic schwa. What is interesting for us is that the [kl] sequence possibly arising from IOG does not sound at all like the /kl/ sequence where /l/ is syllabic, as in word like 'buckle'.

We have now seen that none of the conditions are met for the p-licensing of N2 in (32): it is not domain final, it is not properly governed, it is not magically licensed, nor is it subject to Inter Onset Government. From this we can conclude that N2 must be phonetically realised, and our next question arises: How is this N2 phonetically realised? The results reported in §4 show that words containing a potential syllabic /l/ never contained a schwa, whilst words containing a potential syllabic /n/ either contained no schwa, when the /n/ was preceded by a coronal, or such words contained a schwa-/n/ sequence, when the preceding syllable ended in a non coronal.

To explain this behaviour I propose the following structures for syllabic /l/ and syllabic /n/:

(34) Structures for syllabic /l/ and syllabic /n/



Syllabic /l/, unlike non syllabic /l/ is attached directly to a nuclear constituent, whilst syllabic /n/ is attached to both a onset constituent, and the preceding nuclear constituent, as a result of spreading. Within Government Phonology literature, such a nucleus, which shares its content with a following onset is known as a pseudo-empty nucleus (Hawarth 1994, Charette 1998)

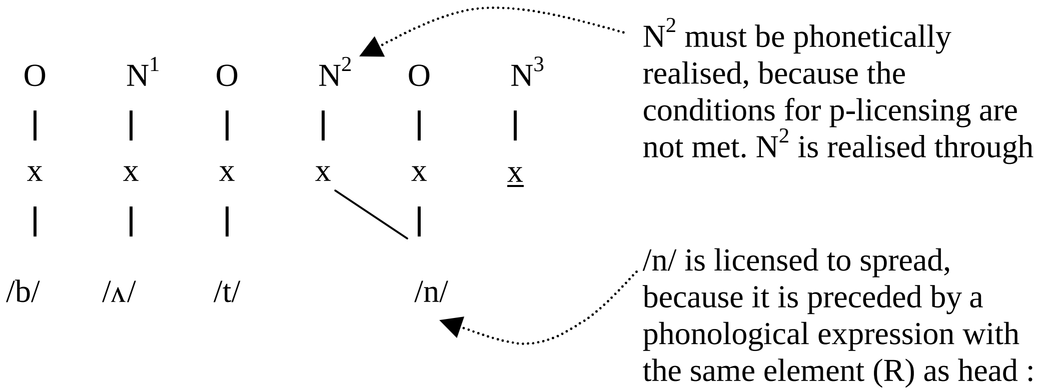
shall maintain the common assumption that (head initial) Inter Onset Government is subject to Constituent-like substantive constraints.

[29] When the /t/ is not normally released, as is often the case with stops preceding syllabic consonants, its elemental representation is (@.ʔ. R. H)

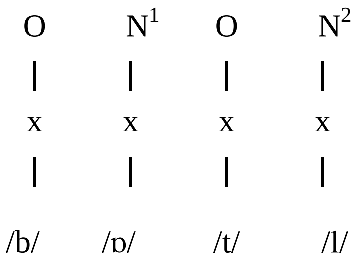
[30] Or rhyme onset sequence, where IOG is head final.

This spreading takes place when the /n/ is Head Licensed by the preceding consonant. Head Licensing is established when two onsets, adjacent at the relevant projection, contain the same element as head. This gives syllabic structures for words like 'button' and 'bottle' as in (35 a,b):

(35a) Syllabic representation of 'button' [bʌtn̩]



(35b) Syllabic representation of 'bottle' [bɒtl̩]

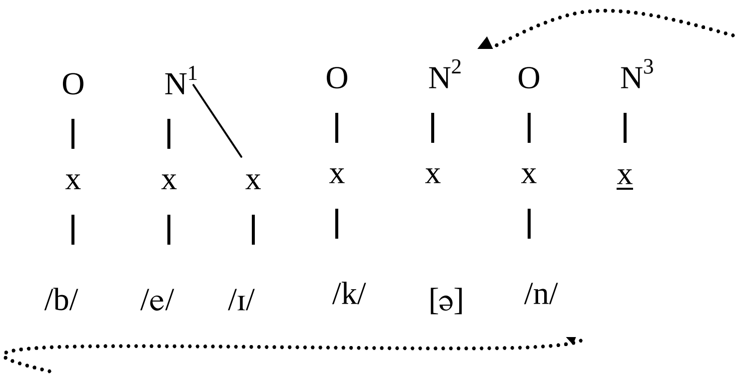


On what grounds do I make these proposals? First it is clear from minimal pairs such as 'gambolling' [gæmbl̩ɪŋ] vs. 'gambling' [gæmblɪŋ], 'finally' [faɪnl̩i] vs. 'finely' [faɪnli], 'evening' (<vb. meaning 'to smooth') [i:vn̩ɪŋ] vs. 'evening' (time of day) [i:vnɪŋ], that syllabic /l/ and syllabic /n/ are not the same as their non syllabic counterparts. This rules out representations whereby syllabic /l/ and syllabic /n/ are directly and uniquely attached to an onset constituent, as is the representation of non-syllabic /l/ and /n/.

The experimental results in §4 showed that for potential syllabic /n/ there is some free variation between [n̩] and [ən], irrespective of context. For example, following /t/, [n̩] occurred 85% of the time, whilst [ən] occurred 15% of the time. Additionally, the suffix '-en', one possible source for syllabic /n/, surfaces either as [n̩], as in [redn̩], or as [ən], as in [swəʊlən], depending on the (coronal) place of the consonant preceding the suffix. The possibility for this variation to exist must somehow be captured in whatever representation is proposed for syllabic /n/.

This variation can indeed be captured by proposing the structure for syllabic /n/ as given in (34), and *in addition*, a condition on the spreading of /n/: /n/ may only spread into the preceding nucleus when head licensed to do so by an onset adjacent at the relevant projection *with the same element as head*. When the onset preceding /n/ is not capable of head licensing the /n/, no spreading takes place. In such circumstances, the intervening nucleus must still be realised, but in these cases it is realised as a default vowel, or in other words, as a schwa.

(36) Syllabic representation of 'bacon'



N^2 must be phonetically realised, because the conditions for p-licensing are not met. N^2 is realised as a

/n/ is not licensed to spread, because it is preceded by a phonological expression which does not have the same element as head: /k/ has (<u>@</u>.ʔ.h.H) as its internal representation. Crucially, this phonological expression is headed by the element @, not the element R. Thus /k/ and /n/ do not share the same head element, and the spreading of /n/ is not licensed.

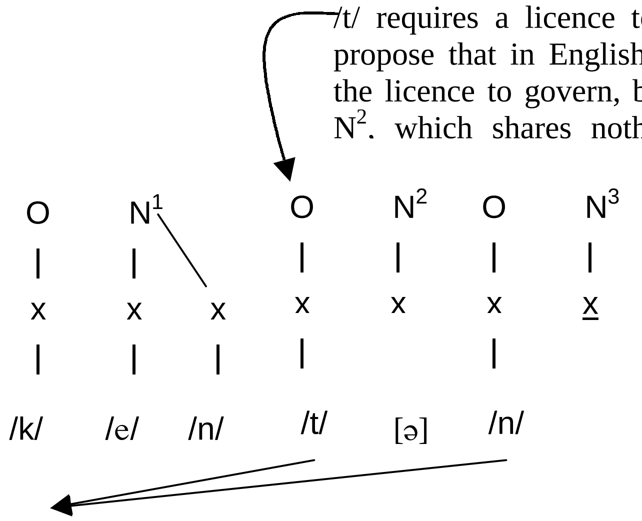
As for why syllabic /n/ is never found following a nasal-stop cluster, even when that nasal-stop cluster is coronal, there are two possible explanations: (1) some sort of OCP effect prevents the spreading of /n/ or (2) when a potential syllabic /n/ is preceded by a cluster, Head Licensing comes into conflict with Government Licensing (in much the same way as we saw Government Licensing coming into conflict with Proper Government in §5.3). In this conflict Government Licensing wins out as, I shall claim, a License to Govern cannot be given by a nucleus with content shared with other constituents. Briefly I consider each of these proposals in turn below.

In the experiment described in §4 I investigated only nasal-stop clusters preceding potential syllabic consonants. Without evidence as to the behaviour of potential syllabics following other types of clusters e.g. /st/ clusters as in 'crystal' or 'Houston', it is not possible to rule out some sort of OCP influence of the /n/ in the cluster on the potential syllabic /n/, forcing a schwa to occur between the two /n/s to separate them. I remain sceptical of this proposal as the two /n/s are not adjacent at any level, given the intervening stop consonant and therefore interaction would be unlikely, but until further experimental evidence is gathered the possibility of an OCP effects has to be entertained.

As an explanation for the non occurrence of syllabic /n/ following a /nt/ cluster I would rather adopt an approach under which two licensing principles (Head Licensing and Government Licensing) come into conflict, the higher ranked of the two (Government Licensing) winning out. Government Licensing is concerned with the ability of a onset head to license its complement. In order for that head to license its onset, it must receive licence to do so. As we saw in §6.3 languages vary as to what may provide the licence to govern to a onset head (Charette 1990, 1991a). In some languages a p-licensed nucleus may act as a Government licensor (e.g. Polish), whilst in others only a realised nucleus, i.e. one with phonetic content, may act as a Government licensor (e.g. French). Thus occasions can arise when Proper Government and Government Licensing come into conflict and the solution depends on the language specific ranking of these two principles. I wish to extend these ideas in two ways (1) to propose that Head Licensing and Government Licensing can also come into conflict in those languages where both principles are active, and (2) to propose that a licence to govern can be given by either a nucleus which shares its content with an adjacent onset, i.e. a pseudo-empty nucleus, or by a nucleus which does not share any content with an adjacent onset. In the latter case, the nucleus may contain either melodic content or be a genuinely empty (as opposed to a pseudo-empty) nucleus. A parameter setting decides which option is chosen. In English, the parameter is set such that a nucleus which provides a license to govern must not be a pseudo-empty nucleus.

(37) The interaction of Head Licensing with Government Licensing: the failure of a syllabic /n/ to be realised after an /nt/ cluster.

/t/ requires a licence to govern the preceding /n/. I propose that in English, a shared PE cannot provide the licence to govern, but that a phonetically realised N^2, which shares nothing, can indeed provide the



/t/ and /n/ both have the element (R) as their head, and thus the conditions for Head Licensing are met. However /n/ does not spread, because the principle of Government Licensing, along with the language specific parameter that precludes a shared PE from providing the licence to govern, outranks the

As for syllabic /l/, we saw in §4 that it *is* found after clusters as well as singletons. This is not surprising if we accept the proposals that syllabic /l/ is attached directly and uniquely to a nucleus. The PE representing syllabic /l/ is not shared, and thus when it comes to Government Licensing the head of the preceding cluster, no conflict arises, and syllabic /l/ may indeed follow the given cluster, unlike syllabic /n/

Reconsidering now to the representation of syllabic /l/, I propose that this segment be directly attached to a nuclear constituent. Theory internally to Government Phonology there are no restriction on the set of elements which may attach to either a nuclear or non-nuclear constituent (cf. the one mouth principle, Anderson and Ewen 1987, Clements 1991). Theory externally, however, we must, as an upshot of my proposal, ask ourselves whether it is legitimate to treat syllabic /l/ as a vowel.

Let us consider definitions which have been made concerning what is and what is not a vowel or a consonant. A vowel can be defined articulatorily as involving no major stricture in the vocal tract (Ladefoged and Maddieson 1996:281), or acoustically as having relatively a long interval of periodic energy with three clear formants (Johnson 1997). Phonologically speaking, a structural definition is usually given for vowels: that which is found at the centre of a syllable, or acts as a syllable on its own. On each of these grounds it could be argued that syllabic /l/ is indeed very vowel-like. Phonetically speaking, /l/ is the most sonorous of oral consonants, having vowel-like formant characteristics, and can indeed be produced without any occlusion (Ladefoged and Maddieson 1996). Syllabic /l/ is also vowel-like phonologically speaking: it may occur after any onset (recall that syllabic /n/ does not show this behaviour), and can indeed form a syllable on its own, as in 'bottle'. Thus there seem to be no strong grounds for ruling out my proposal that syllabic /l/ is indeed attached directly to a nuclear constituent. Such an analysis would explain why no variation is seen in the realisation of potential syllabic /l/ (there is no 'space' in the syllabic structure, contrary to the case of syllabic /n/).

According to my proposals syllabic /l/ is attached to one timing slot whilst syllabic /n/ is attached to two timing slots. In some sense, then, I would appear to be claiming that syllabic /n/ is longer than syllabic /l/. The data presented in (11) and (14) do appear to support the claim that syllabic /l/ and syllabic /n/ have different durational structures although it is clear that my data does not support any claim that syllabic /n/ is twice as long as syllabic /l/; in my data syllabic /n/ is approximately 20% longer than syllabic /l/ (this difference cannot be due to any inherent differences between /l/ and /n/, e.g. because of the different articulators involved, on the grounds that the difference between onset /l/ and onset /n/ is only approximately 6%). This prompts the questions as to how skeletal slots should be phonetically interpreted. An in-depth exploration of this issue is, unfortunately, beyond the scope of the current paper but I shall briefly identify some of the issues that should be borne in mind.

No author, to the best of my knowledge, claims that all timing slots are equal. First of all there are the well known facts concerning onsets and their non-role in syllable weight calculation. GP researchers would not refute these facts, yet they maintain that skeletal positions are in fact attached to onsets. Thus, at some levels, such researchers must recognised that the skeletal slot attached to the [k] of [kat] is not the same as the skeletal slot attached to the [a] or [t] of that same expression. Some researchers working within Moraic Phonology (e.g. Hubbard 1995) also recognise that not all moras are equal, making a distinction between syllabic/head moras (those attached to nuclei) and non-syllabic/weak moras (those attached to the rhyme/onset). Another issue that needs to be considered is the perception of duration. Whilst the difference between syllabic /l/ and coda /l/ or syllabic /n/ and onset /n/ may be statistically significant in the raw time dimension, it is not clear that such differences are perceptually significant. Finally, it may be the case that intrinsic segmental phonetic length is not cognitively represented at all, but is merely a by-product of either the acoustics or the articulation (cf. Fowler 1977), depending upon one's beliefs about the relative importance of these two spheres in phonology.

8 Conclusions

In this paper I have presented the results of an acoustic investigation into the nature of syllabic /l/ and syllabic /n/ in Southern British English. These showed that syllabic /l/ and syllabic /n/ behave quite differently from each other, in at least three ways. The distribution of syllabic /n/, but not syllabic /l/, is sensitive to the place of articulation of the immediately preceding singleton consonant. A potentially syllabic /n/ is only actually realised as syllabic when preceded by a homorganic, i.e. coronal consonant. When /n/ is preceded by a non coronal consonant, the potential syllabic consonant is actually realised as /ən/. Syllabic /l/, on the other hand has no such restrictions on its distribution, for a potential syllabic /l/ is realised as a syllabic /l/ (distinct from a schwa-/l/ sequence) irrespective of the place of the preceding consonant.

Secondly, the distribution of syllabic /n/ is also affected by the structural complexity of its context, i.e. whether or not it is preceded by a singleton or a cluster: if a potential syllabic /n/ is preceded by a nasal-stop cluster, the potential syllabic /n/ is always realised as /ən/, even when preceded by a homorganic cluster. Once again, syllabic /l/ behaves in a different way: a potentially syllabic /l/ is always realised as an actual syllabic /l/ irrespective of the place of the preceding nasal-stop cluster. Thirdly, with respect to duration syllabic /n/ is akin to coda /n/, both being significantly longer than onset /n/. Syllabic /l/, however, is not akin to coda /l/ in duration, but is rather significantly shorter than coda /l/, and approximately the same duration as onset /l/.

In the second half of this paper I proposed that these differences in behaviour can best be explained by proposing differences in their syllabic structure. This is a new proposal, quite unlike those made by previous researchers. Chomsky and Halle, for example, proposed that syllabic /l/ and syllabic /n/ should have fundamentally the same feature bundles:

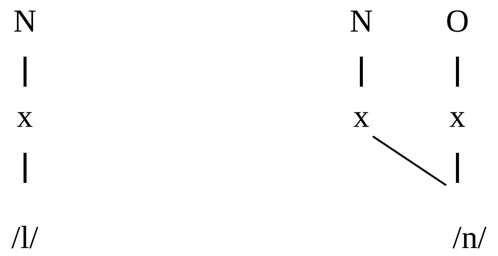
(38) Feature tables from Chomsky and Halle 1968:354

	SONORANT	SYLLABIC	CONSONANTAL
VOWELS	+	+	-
SYLLABIC LIQUIDS	+	+	+
SYLLABIC NASALS	+	+	+
NON SYLLABIC LIQUIDS	+	-	+
NON SYLLABIC NASALS	+	-	+
GLIDES	+	-	-
OBSTRUENTS	-	-	+

Wells (1995), Gimson (1989) and Gussmann (1991) have also proposed that syllabic /l/ and syllabic /n/ be treated uniformly, both being underlying composed of a schwa+consonant sequence.

I, on the other hand, propose that syllabic /n/ and syllabic /l/ be represented differently, on account of their different behaviours. If syllabic /n/ and syllabic /l/ are represented differently it is not surprising that they also behave different. The representations I proposed in (34) are here repeated as (39) for convenience:

(39) Structures for syllabic /l/ and syllabic /n/

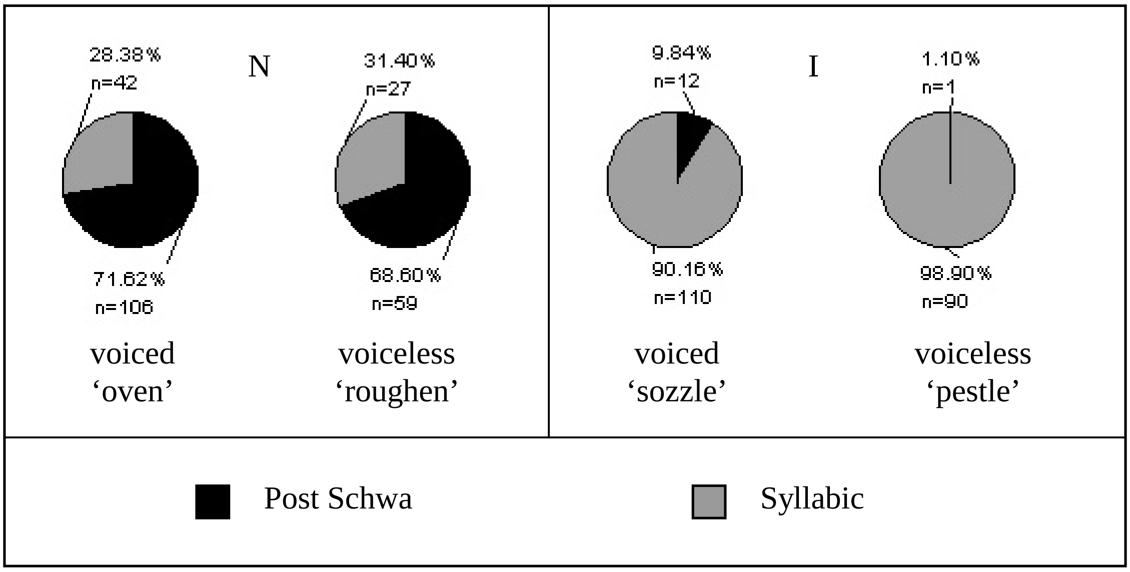


By proposing that syllabic /l/ is attached directly to a nuclear constituent I capture that fact that potential syllabic /l/ is always realised as syllabic /l/ regardless of the nature of the preceding onset. As discussed in §7 above, there are no theoretical or definitional grounds for ruling out the interpretation of syllabic /l/ as a vowel. By proposing that syllabic /n/ is attached to a onset, and that this /n/ may spread when the conditions for head licensing are met, I capture and explain the variability of potential syllabic /n/, which is sometimes realised as [n̩] and sometimes as [ən], when preceded by a singleton. Finally I explain the non occurrence of syllabic /n/ after clusters by proposing that phonological expressions which are shared between two constituents are not able to provide a license to govern to the head of the preceding cluster.

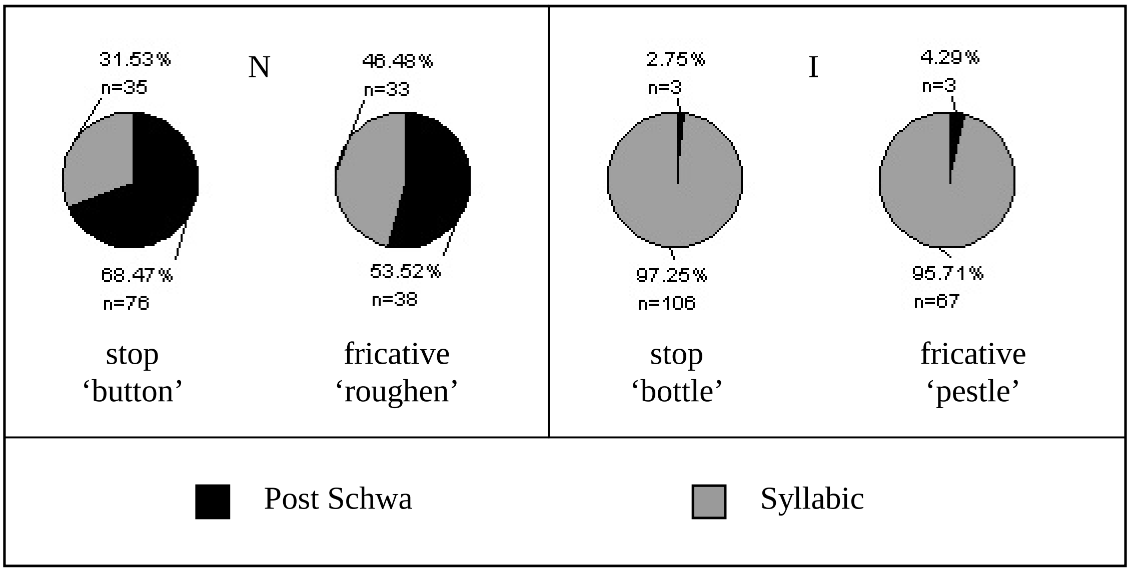
Questions which still remain to be investigated acoustically, and which may provide further answers to help with a phonological analysis include the role of voicing and manner in the distribution of syllabic consonants. Preliminary results from the pilot study of one speaker with an expanded set of singleton contexts suggest that voicing does not effect the distribution of syllabic /n/ (χ^2 (1) = 0.12, p = 0.73), but does effect the distribution of syllabic /l/ (χ^2(1) = 5.50, p = 0.019). In addition these results suggest that the distribution of syllabic /l/ is not affected by the manner of the preceding singleton (χ^2 (1) = 0.017, p = 0.896)[31]. With regard to /n/, whilst statistically speaking there is no significant effect of manner on the occurrence of syllabic /n/ (χ^2 (1) = 3.520, p = 0.061), both visual inspection and the proximity of the p value to the arbitrarily set level of significance suggest that manner may play some role in distribution of syllabic /n/, with more occurrences of syllabic /n/ following fricatives rather than stops. These may provide fruitful ground for future research.

[31] This chi-squared was obtained from a table where 50% of the cells have an expected count less than 5. As a consequence the value obtained for p may be inaccurately low.

(40) Role of voicing in distribution of syllabic /l/ and syllabic /n/ (subject 1)



(41) Role of manner in distribution of syllabic /l/ and syllabic /n/ (subject 1)



In conclusion I have shown that syllabic /l/ and syllabic /n/ behave in ways which do not justify a uniform analysis. Instead I have proposed distinct syllabic structures and new interactions of well established principles in Government Phonology, providing an analysis which I believe can elegantly capture different behaviours of syllabic /l/ and syllabic /n/ in Southern British English. I have pointed out some areas for further research into the same topic and would hope to stimulate investigation into syllabic consonants in other dialects and languages. For example, it would be interesting to investigate syllabics in tapping dialects of English, where syllabic /l/ and syllabic /n/ have different effects on the preceding consonant:

when not part of a cluster, /t/ is flapped before syllabic [l̩], but not before syllabic [n̩] e.g. Am. Eng. [bʌtn̩] but [bɑːt̬l̩] (Wells 2000). I also hope that this paper has shown how phonological insight can be informed by phonetic research.

Acknowledgements

This research was funded by the Students Award Agency for Scotland, Grant No. 10484396. I am very grateful to my informants, audiences at the 9th Manchester Phonology Meeting, ConSOLE X, Monik Charette, Bernard Howard, Wouter Jansen, Jahan Latif and Justin Watkins.

References

Anderson, J. & Ewen, C. (1987) *Principles of Dependency Phonology*. Cambridge: Cambridge University Press.
Árnason, K. (1980) *Quantity in historical phonology: Icelandic and related cases*. Cambridge: Cambridge University Press.
Barnes, J. (2001) Syllable-weight isochrony and speech timing in Turkish. Talk given at the Conference on the Phonetics-Phonology Interface, ZAS, Berlin.
Barry, M. (2000) A phonetic and phonological investigation of English clear and dark syllabic /l/. *Bulletin de la Communication Parlée No. 5* (Les Cahiers de l'I.C.P). pp 77-87.
Beckman, M. (1982) Segment duration and the 'Mora' in Japanese. *Phonetica* 39:113-135.
Bell, A. (1978) Syllabic consonants. In Greenberg, J. (Ed.) *Universals of Human language. Volume 2: Phonology*. Stanford: Stanford University Press. pp 153-201.
Blevins, J. (1995) The syllable in phonological theory. In J. Goldsmith (ed.) *The handbook of phonological theory*. Oxford: Blackwell.
Bloomfield, L. (1933) *Language*. London: George Allen & Unwin Ltd.
Bromberger, S. & Halle, M. (1989) Why Phonology is Different. *Linguistic Inquiry* 20: 51-70.
Byrd, D. (1993) 54,000 American stops. *UCLA Working Papers in Phonetics* 83:97-115.
Charette, M. (1990) License to govern. *Phonology* 7:233-253
Charette, M. (1991a) *Conditions on phonological government*. Cambridge: Cambridge University Press.
Charette, M. (1991b) Mongolian and Polish meet Government Licensing. *SOAS Working Papers in Linguistics and Phonetics* 2:275-291
Charette, M. (1998) Empty and pseudo empty categories. *SOAS Working Papers in Linguistics and Phonetics* 8:167-176.
Chomsky, N. & Halle, M. (1968) *The sound pattern of English.* New York: Harper and Row.
Chomsky, N. (1981) *Lectures on Government and Binding*. Dordecht: Foris Publications.
Chomsky, N. (1982) *Some Concepts and Consequences of the Theory of Government and Binding*. Cambridge, MA: MIT Press.
Clark, J. & C. Yallop (1995) *An introduction to phonetics and phonology*. 2nd edition. Oxford: Blackwell.
Clements, G. (1991) Place of articulation in consonants and vowels: a unified theory. *Working Papers of the Cornell Phonetics Laboratory* 5:77-123
Cowan, W. & Rakusan, J. (1985) *Source book for linguists.* Philadelphia: John Benjamins.
Cruttenden, A. (1994) *Gimson's Pronunciation of English*. 5th edition. London: Arnold.
Crystal, D. (1987) *The Cambridge Encyclopaedia of Language*. Cambridge: Cambridge University Press.
Crystal, T. & House, A. (1987a) Segmental durations in connected speech signals: Current results. *Journal of the Acoustical Society of America* 83:1553-1573.
Crystal, T. & House, A. (1987b) Segmental durations in connected speech signals: Syllabic stress. *Journal of the Acoustical Society of America* 83:1574-1585.
Culicover, P. (1997) *Principles and Parameters: An Introduction to Syntactic Theory*. Oxford: OUP.
Cutler, A., Dahan, D. & van Donselaar, W. (1997) Prosody in the comprehension of spoken language: a literature review. *Language and Speech* 40:141-201.
Cyran, E. (2000) Against licensing of empty nuclei. Handout from talk given at Eötvös Loránd University, Budapest.

Fidelholtz, J. (1975) Word frequency and vowel reduction in English. *Proceedings of CLS* 11:200-213.
Fowler, C. (1977) *Timing control in speech production*. Bloomington: Indiana University Linguistics Club.
Garnes, S. (1976) *Quantity in Icelandic: production and perception*. Hamburg: Buske Verlag.
Gimson, A. (1989) *An introduction to the pronunciation of English*. 4th Edition. London: Edward Arnold.
Gussmann, E. (1991) Schwa and syllabic sonorants in a non-linear phonology of English. *Anglica Wratislaviensia* XVII 1061:25-39.
Gussmann, E. and Cyran, E. (1998) 'Polish consonantal sequences: a phonological testing ground' in Cyran, E. (ed.) *Structure and Interpretation Studies in Phonology*, pp. 127-137.
Harris, James (1983) *Syllable structure and stress in Spanish: A nonlinear analysis*. Cambridge, MA: MIT Press.
Harris, John. (1990) 'Segmental complexity and phonological government' *Phonology* 7.2:255-300.
Harris, J. (1994) *English Sound Structure*. Oxford: Blackwell.
Harris, J. & Lindsey, G. (1995) The elements of phonological representation. In Durand, J. & Katamba F. *Frontiers of Phonology: Atoms, structures, derivations*. pp 34-79.
Harris, J. & Lindsey, G. (2000) Vowel patterns in mind and sound. In Noel Burton-Roberts, Philip Carr & Gerry Docherty (eds.) *Phonological knowledge: Conceptual and empirical issues.* pp. 185-205.
Hawarth, E. (1994) The trouble with French glides. *SOAS Working Papers in Linguistics and Phonetics* 4:53-70.
Heo, Y. (1994) *Empty Categories and Korean Phonology*. PhD. Thesis. School of Oriental and African Studies, University of London, London.
Hubbard, K. (1995) 'Prenasalised consonants' and syllable timing: evidence from Runyambo and Luganda. *Phonology* 12: 235-256.
Ito, J. (1987) Syllable theory in Prosodic Phonology. PhD Dissertation, University of Massachusetts.
Jensen, S. (1994) Is ? and element? Towards a non-segmental phonology. *SOAS Working Papers in Linguistics and Phonetics* 4:71-78
Johnson, K. (1997) Acoustic and Auditory Phonetics. Oxford: Blackwell.
Jones, D. (1959) The use of syllabic and non-syllabic *l* and *n* in Derivatives of English words ending in syllabic *l* and *n*. *Zeitschrift Für Phonetik* 12 (10):249-258.
Kaye, J. (1989) *Phonology: A cognitive view*. Hillsdale, N.J.: Lawrence Erlbaum.
Kaye, J. (1990a) 'Coda' Licensing. *Phonology* 7:301-30.
Kaye, J. (1990b) Government in phonology: the case of Moroccan Arabic. *The Linguistic Review* 6:131-159.
Kaye, J. (1991) 'Do you believe in magic? The story of S + C sequences' *SOAS Working Papers in Linguistics and Phonetics* 2:293-314.
Kaye, J. (1995) Derivations and interfaces. In J. Durand and F. Katamba (eds.) *Frontiers of Phonology*. London: Longman. pp.289-332.
Kaye, J., Lowenstamm, J. and Vergnaud, J-R. (1985) The internal structure of phonological representations: a theory of charm and government'. *Phonology Yearbook* 2:305:28.
Kaye, J., Lowenstamm, J. and Vergnaud, J-R. (1990) Constituent structure and government in phonology. *Phonology* 7(2): 193-231
Ladefoged, P. & Maddieson, I. (1996). *The sounds of the world's languages.* Oxford: Blackwell.
Laver, J. (1994) *Principles of phonetics*. Cambridge: Cambridge University Press
Lee, H-J (1999) Inter-onset government: its needs and interaction with proper government. SOAS Working Papers in Linguistics 9:253-274.
Lehiste, I. (1964) *Acoustical Characteristics of Selected English Consonants*. Indiana University Research Center in Anthropology, Folklore and Linguistics 34. Bloomington: Indiana University.
O'Shaughnessy, D. (1981) A study of French vowel and consonant durations. *Journal of Phonetics* 9:385-406.
Perkell, J., Guenther, F., Lane, H., Matthies, M., Perrier, P., Vick, J. Wilhelms-Tricarico, R., and M. Zandipoor (2000) A Theory of speech motor control and supporting data from speakers with normal hearing and with profound hearing loss. *Journal of Phonetics* 28:233-272
Ploch. S. (1999) Nasals on my mind. PhD Thesis, School of Oriental and African Studies, University of London, London.
Price, P. (1980) Sonority and syllabicity: acoustic correlates of perception. *Phonetica* 37: 327-343.
Rietveld, T. & R. van Hout (1993) *Statistical techniques for the study of language and language behaviour*. Berlin: Mouton de Gruyter.
Roach, P. (1991) *English phonetics and phonology*. 2nd Edition. Cambridge: CUP.
Roach, P. (2000) *A small encyclopaedia of phonetics*. http://www.linguistics.reading.ac.uk/staff/Peter.Roach
Roach, P., Sergeant, P. & Miller, D. (1992) Syllabic consonants at different speaking rates: A problem for automatic speech recognition. *Speech Communication* 11: 475-479
Rockey, D. (1973) *Phonetic Lexicon*. London: Heyden & Son Ltd.
Rubach, J. (1977) Changes of consonants in English and Polish: a generative account. *Prace J zykoznawcze* 87:1-169
Scheer, T. (1998) Governing Domains are Head Final. In Cyran, E. (Ed.) *Structure and Interpretation: Studies in Phonology*. Lublin: Folium. pp 261-285)

Selkirk, E. (1982) The syllable. In H. van der Hulst & N. Smith (eds.) *The structure of phonological representations*, vol. 2. Dordecht: Foris Publications. Reprinted in J. Goldsmith (ed.) (1999*) Phonological Theory: The essential readings*. Oxford: Blackwell.
Spencer, A. (1995) *Phonology: theory and description*. Oxford: Blackwell.
Sproat, R. & Fujimura, O. (1993) Allophonic variation in English /l/ and its implications for phonetic implementation. *Journal of Phonetics* 21: 291-311.
Stevens, K. (1998) *Acoustic Phonetics*. Cambridge, MA: MIT Press.
Umeda, N. (1976) Consonant duration in American English. *Journal of Acoustical Society of America* 61,3:846-858.
Wells. J. (1982) *Accents of English* (3 volumes). Cambridge: Cambridge University Press.
Wells, J. (1995) New syllabic consonants in English. In J. Windsor Lewis (ed.) *Studies in General and English Phonetics*. London: Routledge. pp 401-412.
Wells, J. (2000) *Longman pronunciation dictionary*. 2nd edition. Harlow: Longman.
Williams, G. (1997) A pattern recognition model for the phonetic interpretation of elements. *SOAS Working Papers in Linguistics and Phonetics* 7:275-297.
Woods, A., Fletcher, P. & A. Hughes (1986) *Statistics in language studies*. Cambridge: CUP.
www.praat.org

The Weight of Phonetic Substance in the Structure of Sound Inventories

Nathalie Vallée, Louis-Jean Boë, Jean-Luc Schwartz, Pierre Badin, Christian Abry
Institut de la Communication Parlée, UMR CNRS 5009, INPG – Université Stendhal
BP 25 X – 38040 Grenoble Cédex 9, France

Abstract

In the research field initiated by Lindblom & Liljencrants in 1972, we illustrate the possibility of giving substance to phonology, predicting the structure of phonological systems with *non-phonological* principles, be they listener-oriented (perceptual contrast and stability) or speaker-oriented (articulatory contrast and economy). We proposed for vowel systems the Dispersion-Focalisation Theory (Schwartz et al., 1997b). With the DFT, we can predict vowel systems using two competing perceptual constraints weighted with two parameters, respectively λ and α. The first one aims at increasing auditory distances between vowel spectra (dispersion), the second one aims at increasing the perceptual salience of each spectrum through formant proximities (focalisation). We also introduced new variants based on research in physics - namely, phase space (λ,α) and polymorphism of a given phase, or superstructures in phonological organisations (Vallée et al., 1999) which allow us to generate 85.6% of 342 UPSID systems from 3- to 7-vowel qualities. No similar theory for consonants seems to exist yet. Therefore we present in detail a typology of consonants, and then suggest ways to explain plosive vs. fricative and voiceless vs. voiced consonants predominances by i) comparing them with language acquisition data at the babbling stage and looking at the capacity to acquire relatively different linguistic systems in relation with the main degrees of freedom of the articulators; ii) showing that the places "preferred" for each manner are at least partly conditioned by the morphological constraints that facilitate or complicate, make possible or impossible the needed articulatory gestures, e.g. the complexity of the articulatory control for voicing and the aerodynamics of fricatives. A rather strict coordination between the glottis and the oral constriction is needed to produce acceptable voiced fricatives (Mawass et al., 2000). We determine that the region where the combinations of Ag (glottal area) and Ac (constriction area) values results in a balance between the voice and noise components is indeed very narrow. We thus demonstrate that some of the main tendencies in the phonological vowel and consonant structures of the world's languages can be explained partly by sensorimotor constraints, and argue that actually phonology can take part in a theory of Perception-for-Action-Control.

1 Phonology in a substance-based linguistics

Speech communication operates on two highly-structured levels, the system itself and its physical realisation. This is probably the reason why speech communication is so efficient compared to other communication means used by man or animal. The terms *language* and *speech* refer to these two levels, separated by Saussurean structural linguistics in *form* and *substance*, and reconsidered by generative grammar under the terms *competence* and *performance*. Throughout the 20th century, several axioms of the core of structuralist, and subsequently generativist, approaches have conditioned relationship between phonetics and linguistics:

- the language/speech dichotomy;
- the independence of these two concepts;
- the primacy of language over speech.

This distinction is the result of a particular methodological approach. Linguistics, in order to make empirical data intelligible, separate the study of a sound system – its field of research – from the issue of its physical realisation, which may be variable and polymorphous (Ducrot & Schaeffer, 1995, p. 245). These methodological principles, cumulative in their effects, marginalised any attempt to reveal interactions between the major tendencies observed in the phonological systems of the world's languages – their universals – and the articulatory and acoustic characteristics of their physical realisation. They isolated linguistics in a reductionist internalism and influenced the presuppositions which founded phonology. According to these principles, phonological units cannot be defined by subtantial properties but only with respect to their relative position within the system, and the question of their universality no longer arises. On the threshold of the 21st century, a number of approaches in contemporary linguistics and phonology are still characterised by a strong internalist approach, often presented as an advantage, and goes as far as outright refusal to take into account hypotheses, models, and results obtained by connected disciplines which have language and speech in their field of research. This rejection – to consider the evidence of relationship between form and substance – reiterated throughout this century, is perhaps a unique example in the history of the twentieth-century science: testing data and models, whatever their provenance, should form an intrinsic part of any scientific approach.

In 1952, Jakobson, Fant, and Halle introduced, in *Preliminaries to Speech Analysis* (PSA), a new conception of phonology in linking phonemic features to acoustic correlates and speech perception. Even if their proposed features – which were too general and poorly quantified – have not really clarified the relation between form and substance, the relationship between phonology and phonetics interrupted for almost two decades was discussed anew.

Generative phonology retained from PSA the idea of a universal system of binary features. In *The Sound Pattern of English* (SPE) in 1968, Chomsky and Halle replaced the traditional acoustico-perceptual specification by a universal phonetic representation, expressed in terms of more numerous articulatory features which were precisely defined and well-documented. This can be considered as a very important advance in the framework of phonetic description and relationship between form and substance. It might have been expected that generativist phonologists could have connected their work to articulatory measurements. In fact, the descriptive prolegomena of SPE have remained unfollowed, and the proposed features have not been used in phonological descriptions as a part of a pure symbolic formalism, as the authors themselves stated (Chomsky & Halle, 1968, p. 274). To compensate the lack of naturalism, Chomsky and Halle reintroduced the Theory of Markedness (inherited from the phonology of Trubetzkoy). More recently, the Optimality Theory was proposed to preserve those universal constraints which reveal the unity of language. For Prince and Smolensky (1993), the universal grammar can be essentially considered as a set of ordered constraints, often conflicting, which regulate the well-formedness of representations from which individual grammars are constructed. These constraints are always active, and languages are then distinguished by the way in which conflicts are resolved. Once again, however, it is necessary to proceed with criteria which do not reduce the reasoning to a straightforward tautology.

In fact, the Saussurean dogma taken up by Chomsky (1965): "The classical Saussurean hypothesis of the logical priority of the study of language (and of the generative grammar that describes it) seems almost incontrovertible" has not truly been called into question, except in 1972, with a new perspective that paved the way towards a whole new sweep of research called "substance-oriented" linguistics brought out by The Maximal Dispersion Theory (Liljencrants & Lindblom, 1972) and The Quantal Theory (Stevens, 1972). We place ourselves in these two approaches initiated in the seventies, using recent work on the typology of sound structures (Vallée, 1994 ; Schwartz & al. 1997a ; Stefanuto & Vallée, 1999), and trying to show that some of the tendencies refer to biological constraints on speech production and speech perception human systems, that is, to the substance, not to the form. Our aim is not to refute the existence of a formal phonological level with its intrinsic formal principles and rules, but to try to determine, and if possible quantitatively model, a set of constraints coming from the speech substance and capable of having played a part in the emergence of this formal system – and therefore to throw some lights on phonological facts which could sometimes appear arbitrarily.

Since Lindblom, a number of elements are now available to integrate phonology in a substance-based theory called the Perception-for-Action-Control Theory (Schwartz et al., 2002): *"this theory should be able to show how the choice of speech units inside the phonological system may be constrained and patterned by the inherent limitations and intrinsic properties of the speech perception system – and its indissociable companion, the speech production system"*. The core of the proposal is that *"a listener might follow the vocalisations of his speaking partner, in order perhaps to understand them, but at least certainly to imitate and learn: in other words, perception enables a listener to specify the control of his future actions as a speaker. On the other hand, […] the perceptual representations of speech gestures transform, deform, shape the speaker's gestures in the listener's mind, and hence provide templates that in return also help to specify the control of the speaking partner's own actions."*

This approach is centred on the co-structuring of the perception and action systems in relation with phonology. However, the Perception-for-Action-Control Theory does not fall within the framework of both *"an "auditory" theory in which the sensory-interpretative chain is considered independently of the patterning of sounds by speech gestures, in the search of some "direct link" between sounds and phonemes; and from a "motor" theory […] in which perception is nothing but a mirror of action, in the claim of a direct link between sounds and gestures."*

The studies presented here are at the core of the relationship between phonology and perception for action control. We attempt to show that phonemes, vowels and consonants, are not obviously arbitrary phonological units: phonological systems are in part co-structured by speech perception and action. Considering that phonology should contain the set of formal structures characterising conscious mechanisms for speech control, it is only logical to assume that it is not independent of the ability of the speech production system to produce gestures, and of the speech perception system to recover and shape these gestures. This is the basis both of a theory we have developed for dealing with oral vowel systems, in the line of Lindblom's Dispersion Theory: The Dispersion-Focalisation Theory (DFT) presented in the following section, and of a set of suggested ways to explain plosive vs. fricative and voiceless vs. voiced consonants predominances then developed from our typological analysis based on UPSID phonological systems. In fact, we adopt an epistemological framework using "external" data to phonological description: speech production and speech perception constraints to which it is possible to add some data on ontogenesis and language disturbances (cf. MacNeilage, 1998). Following this approach initiated by Lindblom, models currently permit the prediction of the main tendencies observed in sound systems. It is thus possible to take a close look at the problem of phonological structures and their changes systematically, to establish a precise diagnosis of what can be attributed to speech production/perception, and to list the questions which must be addressed to linguistics and sociolinguistics instead. With such an approach, we do not fall into the trap of the weaknesses of an inductive approach which consists in inferring general laws from isolated observations and can lead us to the error of presupposing the conclusion. We finally illustrate and discuss the inescapable fact that the relationship between phonology and phonetics has to constitute a research field of linguistic sciences.

2 The weight of phonetic substance in vowel structures

2.1 Prediction of the phonological structure of vowel systems: the DFT

2.1.1 General principles

Since the beginning of the 70s, several proposals have been made to predict the phonological structure of vowel systems with *non-phonological* principles, be they listener-oriented (perceptual contrast and stability) or speaker-oriented (articulatory contrast and economy). The so-called "sufficient perceptual contrast" theory (Lindblom, 1986) provides the best global fit with phonological data. However, to overcome its two main problems (that is, the excessive number of high non-peripheral vowels in the model predictions and the impossibility to predict the [i y u] series within the high vowel set), we proposed at ICP a theory based on two principles, that is dispersion and focalisation. These principles specify two basic properties that vowel gestures should have in order to provide a viable sound system for communication. Firstly, gestures should provide sufficiently different acoustic patterns to allow the perception system to be able to recover them without confusions or ambiguities: this is dispersion. Secondly, they should provide salient spectral patterns (formant convergence in vowel spectra), easy to process and characterise in the ear: that is focalisation. While auditory dispersion is a classical concept, focalisation is a principle introduced by ourselves (Schwartz & Escudier, 1987, 1989). The Dispersion-Focalisation Theory (DFT) (Schwartz et al., 1997b) allows us to predict vowel systems through a competition between two perceptual costs: for a given number of vowels, the most frequent system in the world's languages is supposed to be obtained by minimising a global criterion combining a structural dispersion cost and a local focalisation cost.

2.1.2 Implementation

Each vowel is characterised by the formants of its spectrum, that is F_1, F_2, F_3 and F_4, expressed in a perceptual Bark scale. The (F_2, F_3, F_4) set allows to compute an integrated "effective perceptual formant" F'_2. In the Dispersion-Focalisation Theory (DFT), we define a vowel system by a set of vowels in the maximum available formant space and we associate to each system an energy function consisting of the sum of two costs, namely a structural dispersion cost based on inter-vowel perceptual distances – computed through an Euclidean distance in the (F_1, F'_2) space, and favouring large inter-vowel distances – and a local focalisation cost based on intra-vowel perceptual salience, which aims at providing perceptual preference to vowels showing a *convergence between two formants*, that is, vowels with close F_1 and F_2, F_2 and F_3, or F_3 and F_4. The model is controlled by two parameters: λ specifying the weight of F'_2 in respect to F_1 in the dispersion cost, and α specifying the respective weight of the focalisation cost relative to the dispersion cost. Then, for a fixed number of vowels in a system, we implemented various algorithms to select optimal systems, that is systems with the lowest energy (the best compromise of dispersion and focalisation), either locally ("stable systems") or globally ("best systems") (Schwartz et al., 1997b). Our predictions of optimal vowel systems were then systematically compared to vowel inventories, according to the UCLA UPSID Database (Maddieson, 1984 ; Maddieson & Precoda 1989).

2.1.3 Phase spaces

For a given number of vowels, from 3 to 9 (beyond this limit, vowel systems introduce a new dimension, mainly nasality and less often quantity, Vallée, 1994), we can predict, in the DFT framework, different vowel systems in the (λ, α) space. This leads to the determination of what we call "the phase space", a well-known procedure in thermodynamics used to predict the states of a substance (such as the states of water: steam, liquid and ice), as a function of pressure and temperature. The general trend is that, for a given number of vowels in a system, decreasing λ favours peripheral systems while increasing it favours systems with one and then two high non-peripheral vowels; and increasing α favours focal vowels, and particularly stabilises [y] within an [i y u] high series, while this series is unstable when α is set to 0.

Previous work allowed us to verify that these predictions were more or less compatible with the observed preferred phonological vowel systems in the UPSID_{317} database (Maddieson, 1984).

Considering that peripheral systems are generally preferred from 3 to 7 vowels and that the [i y u] series of high vowels exists in a significant amount of cases in the database (about 5% of the cases in the whole database, and 13% of the cases for systems with 7 vowels or more), we showed that setting the λ value around 0.2 - 0.3 and the α value around 0.3 - 0.4 led to quite acceptable predictions (Schwartz et al., 1997b). In the present work, we try to go one step further: we shall attempt to determine where in the phase spaces one can find the different systems, *preferred or not*, existing in UPSID_{451}, and what kinds of "superstructures" can be derived from this analysis.

2.1.4 Structural symmetries between vowel systems: a typological equivalence criterion

2.1.4.1 Prototypical structures in phase spaces

Our previous simulations led to "prototypical systems". These are winning n-vowels systems in the DFT framework, in the sense that they have a minimal global Dispersion-Focalisation (DF) energy, according to the values of the two free parameters λ and α.

We have focused our study on values of n from 3 to 7 because they allow us to capture the most significant phonological tendencies of the UPSID database. The DFT simulation results are given in Figures 2-6, respectively for n = 3, 4, 5, 6 and 7. For each value of n, the phase space determines regions in the (λ, α) space in which a given system wins (with its vowel qualities displayed as black points on a prototypical grid). We see that there are two prototypical systems for n = 3, which we call S_3T_1 and S_3T_2. There are four prototypical systems for n = 4, 5, 6, and five prototypical systems for n = 7; let us call them S_nT_i, with n from 3 to 7, and i from 1 to 5. The global trend is that increasing n increases the dispersion cost of peripheral systems, hence it decreases the λ boundary necessary for making these systems optimal. Hence peripheral systems are favoured with small values of λ. When λ is too small, the vowel space is completely vertically stretched (since higher formants play a minimal part in the determination of vowel phonetic quality); this favours asymmetrical peripheral configurations because of the interactions between front and back peripheral vowels in the systems. Non-peripheral configurations, that is systems with more than two high vowels, appear with large λ values, and when α increases, focal vowels (especially [i] with close F_3 and F_4, other front unrounded vowels together with [y], all with close F_2 and F_3, and back rounded vowels, with close F_1 and F_2) are favoured. Decreasing α leads to replacement of the high rounded vowel [y] with a high vowel acoustically more central (namely [ɨ] or [ɯ]).

2.1.4.2 Reverse prototypical structures

We hypothesised that two structures having the same number of peripheral vowels but systematically replacing front unrounded vowels by back rounded ones with the same height, and vice-versa, are equivalent structures in the sense of DFT, that is to say that they have roughly the same DF energy for a given value of n and of the (λ, α) pair. This was systematically verified by comparing the energy of the S_nT_i prototypical systems with reverse systems that we called S_nT_i*. For example, for n = 4 we compared S_4T_1 = [i o ɛ a] with S_4T_1* = [i e ɔ a], and S_4T_2 = [i ˈeˈ a u] with S_4T_2* = [i ˈoˈ a u], S_4T_3 = [i y u a] and S_4T_4 = [i ɨ u a] having no reverse counterpart. Indeed, we confirmed that S_nT_i* structures have a DF energy quite close to the S_nT_i ones whatever the region of the phase space, that is to say whatever the λ and α values. Pushing the analogy with physics one step further, this reminds us of the "polymorphism" of a number of solids (e.g. metals, or crystals). In this situation, while fusion produces a homogeneous liquid phase, solidification leads to mixtures of two or more variants of the solid phase, all stable and more or less with the same energy. This is exactly what happens here with the two variants within a given phase. Hence, our typologies of phase spaces involve "superstructures" grouping prototypical structures and reverse ones (displayed with white points instead of black ones in Figures 2-6). The relevance of these superstructures for describing the UPSID database will now be discussed in the next section.

2.1.5 Comparing UPSID data with DFT simulations

2.1.5.1 UPSID data reanalysed

Since Trubetzkoy and his *Principles* (1939), taxonomy has not only been an approach of historical linguistics: associated to research on synchronic trends, it constitutes today a main stage in linguistic theories. Institutionalised in 1961, under the aegis of the *Social Science Research Council* during the New-York *Conference on Language Universals*, this research field aims at finding common basic structures in languages – in diachrony as well as in synchrony. *The Language Universals Project* (1967-1976) led to the building up of the *Stanford Phonology Archives* (Greenberg et al., 1978), with which many important studies dealing with typological classification and phonological tendencies, were achieved (Sedlak, 1969; Crothers, 1978; Maddieson, 1984; Vallée, 1994). But all these studies present variegated contents: data are constantly enriched, questions on the materials vary from one author to the other. The UPSID (UCLA Phonological Segment Inventory Database) (317 then 451 languages) gathers phonological systems of languages in the world, sampling more or less uniformly all linguistic families. $UPSID_{317}$ (Maddieson, 1984), then $UPSID_{451}$ (Maddieson & Precoda, 1989) were chosen to approximate a properly constructed quota sample on a genetic database of the world's existing languages.

UPSID was implemented at ICP several years ago and we have been using it for vowel and consonant research. In order to test our hypotheses, we have reanalysed the UPSID database of vowel systems, thanks to a two-step methodology.

The languages in UPSID have 3 to 28 vowels. Firstly, from raw data, that is to say without any typological equivalencies, we obtain 252 types of phonological structures from 3 to 17 vowel qualities. What we call vowel qualities corresponds to "basic segments" (vs. "elaborated" and "complex" segments) in the sense of Lindblom & Maddieson (1988). We note that more than 96% of the languages have from 3 to 10 basic vowel qualities, and if we focus our study on systems with 3 to 7 qualities, we obtain 77% of the 451 languages (348 systems). This is due to the fact that there are in many cases more vowels than vowel qualities in a given system; for instance / i e a o u / is the phonological structure of four UPSID languages of which three have more than 5 vowels: Chipewyan with 14 vowels /i e a o u iː aː uː ĩ ã ũ ĩː ãː ũː /, Siriono (12) / i ĩ e ẽ a ã o õ u ũ iˠ ĩˠ /, Tamang (10) / i e a o u i̤ e̤ a̤ o̤ ṳ /. The systems with nasal, laryngeal, pharyngeal or retroflexed vowels sharing no vowel qualities with a basic segment, as opposed to the systems quoted above, have been discarded for follow-up analyses. These results in eliminating less than 3% of UPSID's languages and 3.4% of languages having from 3 to 7 vowels, that is seven languages with 6 vowels qualities and five languages with 7 qualities, for instance the Cherokee system / i ˈeˈ ˈəˈ a ˈoˈ u / or the Tarascan system /i ɨ˞ ɛ a o u /. At this stage we retain 336 systems of the database.

Secondly, we take into account the so-called "transparency rule" (Schwartz et al., 1997a). This rule states that schwa should be conceived as a separate class, considering that it does not seem to interfere with the other vowels in a system: indeed, schwa added or removed from a system does not disrupt the structural organisation of this system. The "transparency rule" concerns 64 languages from 4 to 8 vowel qualities. For instance we have classed the Ivatan structure / i ˈəˈ a u / as S_3T_2, Achumawi / i ˈeˈ ˈəˈ a ˈoˈ u / as S_5T_2, Ndut / i e ɛ ə a ɔ o u / as S_6T_1, and Fur / i e ɛ ˈəˈ a ɔ o u / as S_7T_2. The "transparency rule" results in slightly increasing the number of systems in the analysis, thanks to six 8-vowels systems which become 7-vowels ones. Hence at this last stage we stay with 342 systems that is to say 75.8% of the database (Figure 1).

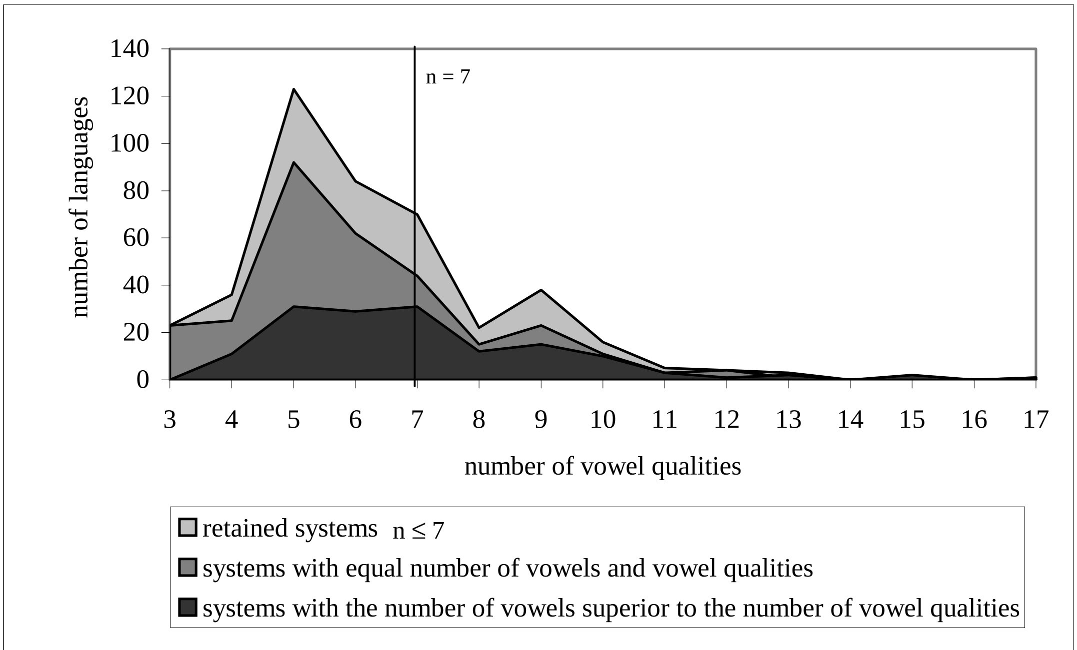


Figure 1: UPSID's languages distribution by number of basic vowel qualities.
We focus our study on systems from 3 to 7 qualities.

2.1.5.2 Distribution of UPSID data within phase spaces

We now have at our disposal both a series of predictions organised around the typology S_nT_i / S_nT_i* defined in Section 2.1.4, and an inventory of 342 systems (the three-fourths of $UPSID_{451}$) with 3 to 7 vowel qualities. The final goal of this work was to try to associate to most of these 342 systems a region in the phase space where they would be optimal (i.e. viable in the sense of the DFT). This is displayed in Figures 2-6, where we have plotted within each region of the phase spaces the number of systems fitting with the corresponding structure. Let us now discuss the obtained results in more detail.

To begin with, it appears that 303 of the 342 3-to-7 vowel systems (88.6%, or 67.2% of the whole $UPSID_{451}$ database) fit with one of the S_nT_i or S_nT_i* types. The 39 rejected systems (fitting with no prototypical or reverse type) correspond to one system in S_3, 5 in S_4, 9 in S_5, 10 in S_6, 14 in S_7; hence their number increases with n, which is logical since the complexity of the distribution of vowel qualities increases.

Next, the most widespread types in Figures 2-6 are those corresponding to S_nT_2 (and sometimes to S_nT_2*). This provides a first confirmation on $UPSID_{451}$ of our results on $UPSID_{317}$ (Vallée, 1994), namely that the (λ, α) region defined by $0.2 \leq \lambda \leq 0.3$ and $0 \leq \alpha \leq 0.4$ is compatible with preferred 3-to-7 vowels systems in UPSID. But the data in Figures 2-6 provide some new confirmation of this result. Indeed, it appears that systems corresponding to types associated to large λ or α values are quite few. On the contrary, most systems on Figures 2-6 are located at low λ values. Indeed, apart from the "best" S_nT_2 structures, other structures generally occupy nearby regions (mostly of types S_nT_1 or S_nT_1*) and if we define a broad acceptable region such as $0.1 \leq \lambda \leq 0.3$ and $0 \leq \alpha \leq 0.4$, we obtain a total of 293 systems, that is 85.6% of the 342 systems of our inventory, which is quite important. Altogether, this confirms with a strong reliability the need to "stretch" the acoustic space along the F_1 dimension in auditory spectral distances, which indicates the dominant role played by the lower formant F_1 in vowels' phonetic quality.

The next observation deals with the symmetry between front and back peripheral vowels. Globally, the data in Figures 2-6 confirm the well-known fact that vowel systems "prefer" both peripheral vowels and front-back symmetry. In the asymmetrical cases, when the numbers of front and back vowels are different, the (classical) trend is that there are more front than back ones: for example, 9 S_4T_1 [i o ɛ a] vs. 0 S_4T_1* [u e ɔ a] structures, 8 S_4T_2 [i u ˈeˈ a] vs. 4 S_4T_2* [i u ˈoˈ a] structures, 14 S_6T_1 [i u e ˈoˈ ɛ a] vs. 10 S_6T_1* [i u o ˈeˈ ɔ a] structures. When the number of front and back vowels are the same, the (less classical) trend is that front vowels have often a more open degree than back ones; though this is not true for 3-vowels systems (2 S_3T_1 [i ˈoˈ a] vs. 0 S_3T_1* [u ˈeˈ a] structures), it is clearly the case for 5-vowels systems (4 S_5T_1 vs. 16 S_5T_1*) and for 7-vowels systems (0 S_7T_1 vs. 2 S_7T_1*).

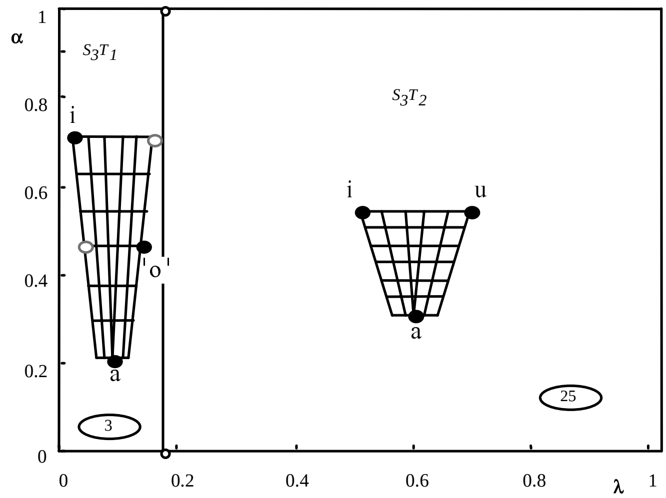


Figure 2: Phase space for three-vowels systems.
S_nT_i structures have their vowel qualities displayed as black dots on a prototypical grid. S_nT_i* vowels are displayed as white dots replacing the black ones. The number in the oval is the total number of UPSID languages with the S_nT_i and S_nT_i*.

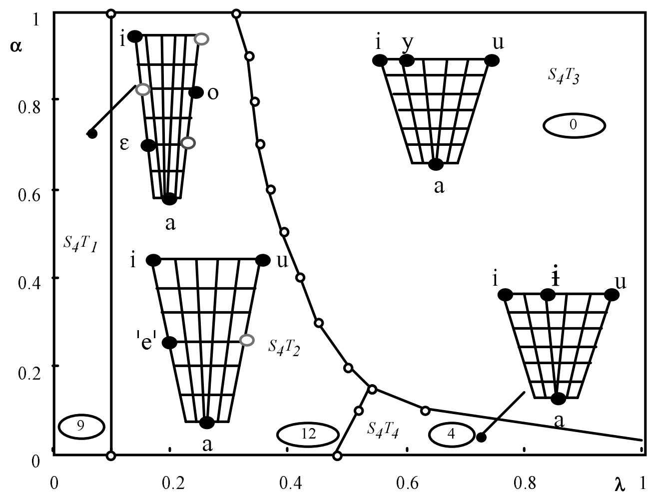


Figure 3: Phase space for the four-vowels systems.

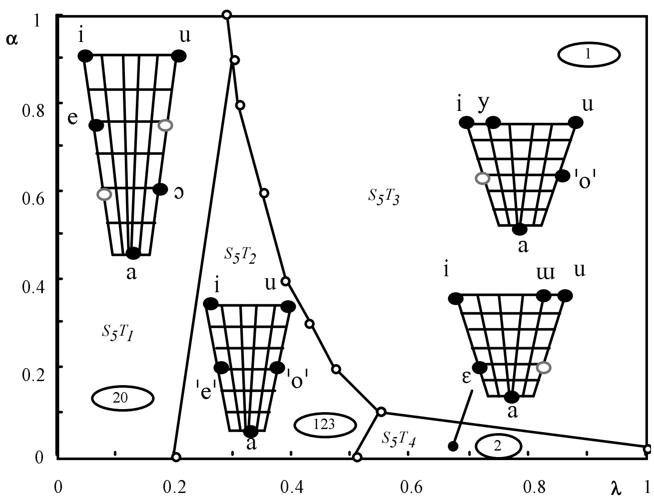


Figure 4: Phase space for the five-vowels systems.

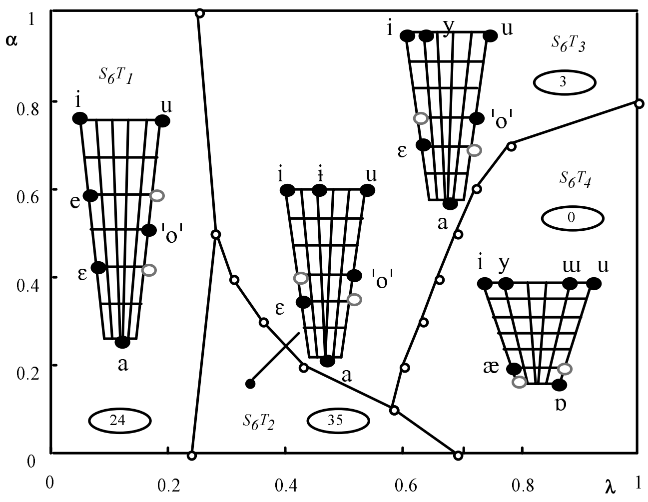


Figure 5: Phase space for the six-vowels systems.

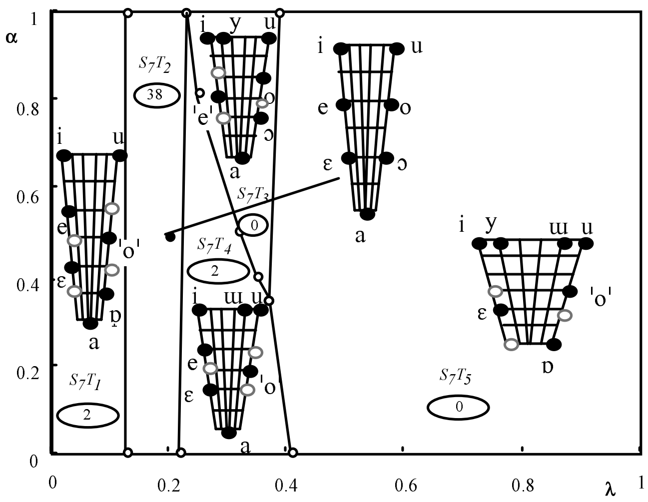


Figure 6: Phase space for the seven-vowels systems.

In what concerns focalisation, its role is more important for stabilising the vowel structures of larger systems, particularly those containing [y]: this vowel appears in only 2% of UPSID's languages from 3 to 7 vowel qualities, but almost 7% of all languages; notice that more than half of them are Indo-European and Uralo-Altaic languages (Alcantara, 1998). In our simulations, /y/ is only present in the S_6T_3 [i y u ˈoˈ ɛ a] structure (3 examples, that is only 1% of our reduced database).

3 Consonant systems and some substantial constraints on form

Whereas we have clearly established a theory for the prediction of vowel systems, no similar theory for consonants exists at present. Therefore we will first describe in detail a typology of consonants, and then suggest ways to explain some of the observed tendencies.

3.1 Typology

3.1.1 Taxonomic elements

Since the 1960’s, typological works have dealt more with vowels than with consonants. This is partly due to the wide range of consonant inventory sizes and the large number of classification parameters that make type emergence difficult. Although the nature of some “external” constraints influencing the content of vowel systems is coming to light (Schwartz et al., 1997a), we are still far from understanding the biological constraints influencing the consonant structures of the world’s languages.

Investigating existing languages is the traditional way to search for universal tendencies, but we also need to look at the question of the capacity to acquire relatively different linguistic systems. Firstly we present a set of consonant inventory tendencies, then we compare them to available language acquisition data – more precisely, at the babbling stage. Indeed, among the different ontogenetic stages, babbling seems to be essential since it marks “precursors” of universal speech attributes, following MacNeilage’s Frame then Content Theory (MacNeilage & Davis, 1990; MacNeilage, 1998).

Basically, our typological study is built on places and manners of articulation, two very usual parameters in consonant classification. $UPSID_{451}$ lists 920 phonemes with 654 consonant segments distributed across 13 places (Figure 7) and 16 manners.

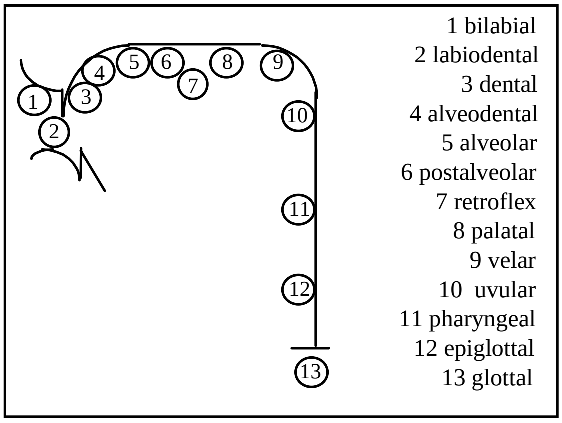


Figure 7: Consonant articulation places ($UPSID_{451}$).

In our typology, manners have been grouped into 7 types: plosives (with implosives, ejectives, glottal stops); nasals; fricatives (with ejectives and *h-sounds*); affricates (with ejectives); approximants; trills/taps/flaps; clicks.

Compared to UPSID, our place classification, based on Creissels (1994) and Ladefoged & Maddieson (1996) is more detailed (17 simple articulation places and 5 double articulation places).

We have kept the retroflex category as a place, as suggested in Maddieson (1984), Ladefoged & Maddieson (1996) and the IPA (1996), though it refers to a manner: the distribution of retroflex consonants across their different places of articulation (alveodental, alveolar, prepalatal, etc.) would require an investigation of the sources of transcription.

We have to stress that making the comparison to previous taxonomies (let us mention works by Hockett (1955), Hagège (1982), Maddieson (1984), Lindblom & Maddieson (1988), Laver (1994)) is not easy in so far as they all differ in their data preparation and classification methodologies.

3.1.2 Distribution

$UPSID_{451}$ languages most often use from 18 to 25 consonants: minimum 6 for Rokotas (a Papuan language, 11 phonemes), maximum 95 (with 48 clicks) for !Xũ (Khoisan family, 141 phonemes). The average size of consonant inventories is 22, with 7.8 plosives, 4.1 fricatives, 3.3 nasals, 2.9 approximants, 2 affricates, 0.6 ejective, 0.5 trill/flap, and 0.2 click.

Table I shows the distribution of consonants across the different places of articulation. Grouping all manners, alveodentals are the most numerous (15.3%), followed by bilabials (14.3%), and velars (12.6%). It is important to note that the coronal class (*i.e.* dental, alveodental, alveolar, postalveolar and retroflex consonants grouped together) represents far and away the largest class of segments, with 44.5% in $UPSID_{317}$, far ahead of bilabials. Though the coronal class represents a wide range of articulation types (Ladefoged & Maddieson, 1996), its use in our typology is acceptable: in inventories, coronal place features seldom serve as phonological distinctions between consonants of this type, except for affricates, and never for anterior coronals (Keating, 1990).

The $UPSID_{317}$ consonant distribution shows that some manners are more represented than others: plosives 38.6%, fricatives 20.2%, nasals 14.6%, approximants 13%, affricates 9.6%, trills/taps/flaps 3.9%. Voiceless oral coronal plosives (type /t/) are the most frequent (they are present in 97.5% of languages); bilabial and coronal nasals (types /m/ and /n/) exist in more than 9 languages out of 10; /k/, /j/, /p/ exist in more than 80% of languages; /w/ and /s/ in 2 out of 3 languages; /d/, /b/, /h/, in more than 60% of languages; one language in 2 has /l/, /g/, /ŋ/, /ʔ/.

alveodental	15.3%
bilabial	14.3%
velar	12.6%
palatal	10%
apico-alveolar	9.9%
glottal	7.6%
lamino-postalveolar	6.7%
labiovelar	5.1%
dental	5%
lamino-alveolar	3.4%
labiodental	3.3%
retroflex	2.9%
uvular	1.7%
others	<1%

Table I: Frequency of occurrences for consonant articulation places from $UPSID_{317}$ (Stefanuto, 1996).

3.1.3 Systems

3.1.3.1 Size of articulation place systems per articulation manners

If the study of the size of consonant inventories gives very little information about their content, contrary to vowels (Schwartz et al., 1997a), the development of a place/manner typology reveals a strong correlation between (i) the size of place systems and (ii) the different articulation manners (Table II):

- In the majority of languages, plosives and nasals occupy, respectively, from 3 to 6 and from 1 to 6 places. The plosive manner is the only one that sound systems universally exploit. Moreover, it is also the only one that systematically uses place contrasts: languages distinguish at least 3 of them. Compared with fricatives, stops (including nasals) are almost always more numerous (see Figure 8).
- Fricatives and approximants are the consonant classes which may recruit the greatest number of place contrasts: fricatives are more widely distributed on the different system sizes. Fricatives have systems of contrasts often more complex than plosives. 8% of languages do not involve fricatives at all (contrarily to plosives, which are universally used).
- For trills, flaps and taps, systems hardly ever present a place contrast – and never spread over more than 2 places. We find the same trends in affricates.

More broadly speaking:

- Languages systematically use place contrasts for plosive consonants; almost one language in two distributes its plosives in a 4-places system.
- Place contrasts are rare for affricates, and exceptional for trills/taps/flaps.
- Approximants frequently spread among 3 articulation places.
- Fricatives spread more widely across the different sizes of place systems.

Places number	1	2	3	4	5	6	7	8
trills, flaps, taps %	**67.2**	4.7						
Affricates %	**44.8**	20.8	6.9	0.3				
Plosives %			28.4	**43.2**	22.4	5.9		
Nasals %	2.5	**31.5**	30.6	25.9	4.5	2.2		
Approximants %	8.2	26.5	**49.8**	8.2	2.5	0.3	0.3	
Fricatives %	3.4	18.3	**25.9**	21.1	14.2	7.3	1.9	0.9
Place opposition Number	-	1	3	6	10	15	21	28

Table II: Size of the place systems in $UPSID_{317}$.
The number of place oppositions changes according to manner; from 0 (*i.e.* 1 place per system) to 28 (8 places). The presented values do not include the voicing feature (voiced and voiceless categories are grouped). 43.2% of the 317 languages have a 4-places plosive system, and 49.8% have approximants spreading on 3 places.

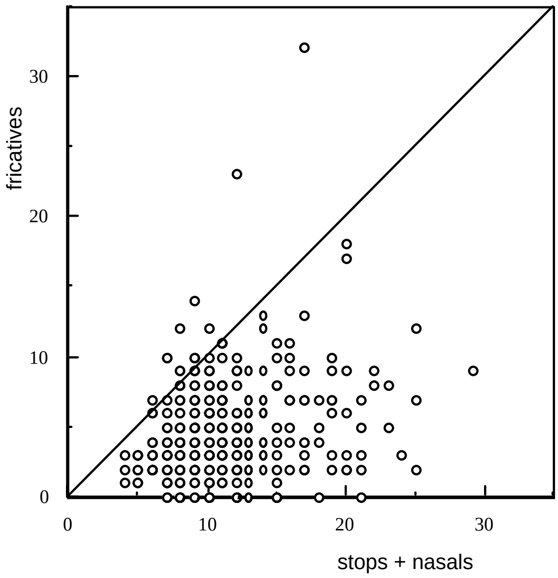


Figure 8: Quantity of fricatives regarding quantity of stops plus nasals in $UPSID_{451}$'s phonological systems.
The line corresponds to languages with an equal number of stops plus nasals and fricatives.

3.1.3.2 Place systems for each manner

Table III reports the content of place systems for each manner. It presents the spreading of plosives, nasals, fricatives and affricates according to the number of articulation places. We observe the following tendencies:

- There is neither a phonological system without plosives, nor a system with only one. When a system distributes the plosives across 2 places, these places are coronal and bilabial. Among 3 places, they are coronal, bilabial and velar. The glottal place appears in 4-places systems. Systems with 5 places recruit bilabial, coronal, velar, glottal and uvular consonants.
- If there is only one nasal in a system, it is a coronal one. If the system has 2 nasal consonants, they are bilabial and coronal. A velar nasal appears in 3 place systems. In larger systems (4 places), the palatal articulation emerges.
- When the fricatives of a system use only one articulation place, they are alveolar, i.e. coronal. When 2 places are used, they are coronal (alveolar) and glottal. In 3 place systems, labiodentals appear; in 4 place systems, the palatal area is exploited, and in 5 ones, velar fricatives emerge.
- Affricates use coronal articulation areas, whatever the place system size (from 1 to 3).

	0 place	1 place	2 places	3 places	4 places	5 places
plosive	0% -	0% -	0.2% coronal bilabial	31% coronal bilabial velar	**43%** **coronal** **bilabial** **velar** **glottal**	21% coronal bilabial velar glottal uvular
nasal	3.5% -	1.8% coronal	31% coronal bilabial	**32%** **coronal** **bilabial** **velar**	27% coronal bilabial velar palatal	3,5% coronal bilabial velar palatal coronal 2
fricative	7% -	6.6% coronal	18% coronal glottal	**28%** **coronal** **glottal** **labiodental**	20% coronal glottal labiodental coronal 2	11% coronal glottal labiodental coronal 2 velar
affricate	33% -	**42%** **coronal**	20% coronal coronal 2	4.9% coronal coronal 2 coronal 3	- -	- -

Table III: Consonant categories (plosives, nasals, fricatives, affricates) from $UPSID_{451}$ are classified according to the number of articulation places, with the percentage of corresponding languages (e.g., 31% of 451 languages present a plosive system spread on 3 places, etc.; 0 place means an absent category). The small differencies observed in the percentages between Table II and Table III come from the number and the choice of languages between the two versions of UPSID. Only the dominant systems are noted under the percentage values.

From these place system distributions per category, we can retain the following main tendencies:

- Generally, an n-element system contains a system with n–1 consonant places,
- When there is only one place, whatever the category (plosive, fricative…), coronal consonants are at stake,
- One can mention a possible link between the major tendencies of the place systems and the degree of articulatory “complexity” of the segments according to Maddieson (1984) and Laver (1994): “complex” segments (e.g. affricates, trills/taps/flaps) use almost exclusively the coronal area whereas “simple” segments use varied places.
- Grouping the most frequent manners, languages prefer contrast with 3 places: more than 1 out of 3 for plosives and nasals and almost 1 out of 3 for fricatives.
- A particular system size dominates each category: 43% of the languages have 4 places for the plosives and 74% have 3 or 4 places; 32% of the languages distribute their nasals on 3 places (59% on 3 and 4); 28% present 3 places for fricatives (48% on 3 and 4); 43% have only one articulation place for affricates.
- Combinations with 5 places exist in only 36% of the languages, mainly for oral plosives (97 languages).
- Combinations with 6, 7 or 8 places are not widely used (15% of the languages). We essentially find them in African systems.

Our results (i) confirm that plosives are far and away the “star consonants”, ubiquitous in languages, and (ii) bring to light trends that appear very regular.

3.1.3.3 Voiced/voiceless ratio

Within the oral plosive category, voiceless sounds (64%) are a lot more numerous than voiced ones (Figure 9). Whatever the place of articulation, the voiceless feature concerns 72% of fricatives (Figure 10). Voiceless affricates are also dominant (74%).

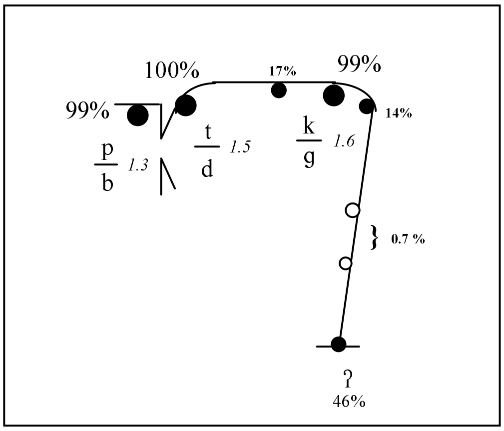


Figure 9: Percentage of UPSID$_{451}$'s languages which investigate the different places of articulation for oral plosives. The voiceless/voiced ratio is mentioned for the most frequent places of articulation. We observe a raising ratio from front to back places of articulation.

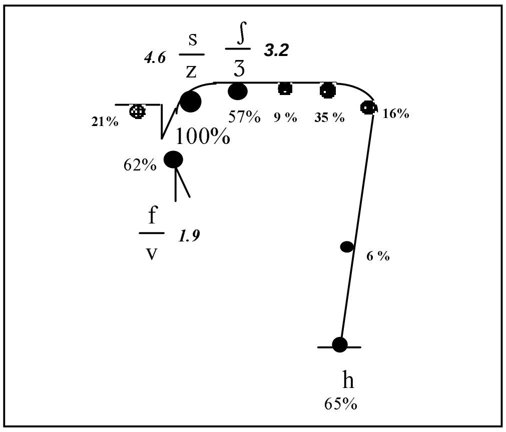


Figure 10: Percentage of UPSID$_{451}$'s languages which investigate the different places of articulation for fricatives. The voiceless/voiced ratio is mentioned for the most frequent places of articulation. The predominence of voiceless fricatives is more important for alveolars and palatals.

3.2 Universal trends and ontogenesis

We gathered the results of several studies on consonant content of production at different stages of canonical babbling (repetition of identical syllables) or variegated babbling (with different types of closures) – the latter being contemporary with the former (Vihman, 1996). All of these studies cover a period from 6-10 (Boysson-Bardies, 1996) to 15-24 months (Stoel-Gammon, 1985), for 15 languages including American English and French. Boysson-Bardies found that [p b m t d n k g ŋ] constitute 80% of consonant-like productions during the first months of babbling, [p b m t d] being the most frequent. The review of papers by Locke (1983) gives, by descending order of frequency, the inventory of consonants representing 95% of occurrences produced by children aged from 6 to 15 months of age: [b] and [m], [p], [d], [h] and [n], [t], [g] and [k], [j] and [w], [s]. Robb & Bleile's observations (1994) during a longer period (8 to 25 months of age) show that, for children aged from 8 to 12 months, several types of closures co-emerge. The most frequent productions are the plosives (oral and nasal) in alveolar and bilabial places. They thus confirm Davis & MacNeilage (1994), who showed that during 7-12 months of age, these consonant-like sounds cover 84% of infants' utterances ([t] [d] 46%, [n] 18%, [p] [b] 15%, [m] 5%); velars and glottals being the last to appear. Some of this experimental evidence has also been corroborated by simulations using linear midsagittal articulatory models developed for three different subjects: Vilain et al. (1999) found that pure mandibular raising, all other articulatory commands being frozen, leads to bi-labial, labio-dental or alveopalatal contacts, but never to velar contacts (also Abry et al., 1997).

If one draws a parallel between the results on language acquisition and typological data, one finds that the types of closures produced during the first months of babbling (8-12 months of age), as well as sounds acquired later during the second year, correspond to the consonant phonemes most frequently used in languages (apart from the order of emergence). One also finds that the preferred place systems, as shown for each manner in our typological analysis, are dominant in the production of babbling: alveolar and bilabial before velar and glottal in the plosive manner (oral or nasal); alveolar and glottal in the fricative manner; labiovelar, coronal and palatal in approximants ([l] and [j] being acquired later, at around 13-16 months of age (Robb & Bleile, 1994). All in all, we thus note that the consonant-like sounds produced during babbling's mandibular oscillations – the Frame, MacNeilage (1998) – correspond to the most frequent consonant inventories of the world's languages – in particular, they tightly correspond to languages with small inventories (Stefanuto, 1997). These findings allow us to form the hypothesis that the more frequent consonants in languages are, most probably, drawn from the stock of potential abilities of babbling. From an articulatory point of view, is it because they are produced with the greatest of ease ? Locke & Pearson (1992, p. 26) did not hesitate to conclude: *"Infants heavily favor stop consonants over fricatives, and there are languages that have stops and no fricatives but no languages that exemplify the reverse pattern. [Such] "phonologically universal" patterns, which cut across languages and speakers are, in fact, the phonetic properties of Homo sapiens."* Ontogenesis thus clearly contributes to supplying some determinating elements in the dependencies between certain general tendencies of phonological systems and the sensorimotor abilities of *Homo loquens.*

3.3 Hints towards explaining main tendencies

3.3.1 Plosive vs. fricative consonants' predominance

Plosive consonants are characterised by the occurrence of a complete occlusion of the vocal tract somewhere between the glottis and the lips (Ladefoged & Maddieson, 1996; Crystal, 1997; Stevens, 1998). The main mechanism underlying the production of plosives is the generation of an impulsive acoustic source associated with the sudden release of the occlusion (Pelorson, 1997). From the control point of view, the main requirement for the articulators is thus to ensure a proper occlusion: this might be achieved by aiming at a virtual articulatory target that would be located beyond the actual vocal tract boundaries. For instance, in the case of the production of an apico-dental plosive, the tongue tip could be assigned a target that would be located slightly above the surface of the hard palate in the alveolar region: this would result in a full contact between tongue and palate, and thus ensure the complete occlusion of the vocal tract. This strategy has successfully been implemented for plosives (Bailly, 1997). It is important to note that this type of strategy does not require a precise control of the articulators: in a first approximation, a broadly defined target, as far as it is set beyond the physical limit of the involved vocal tract boundary, can ensure the realisation of the occlusion.

The situation of fricative consonants is quite different. Fricatives are characterised by the presence of a constriction somewhere in the vocal tract. This constriction induces a local acceleration of the air particle velocity that in turn induces, in conjunction with the shape of vocal tract, turbulence, and thus produces acoustic frication noise sources (Shadle, 1990; Badin et al., 1995; Stevens, 1998; Mawass et al., 2000). The generation of noise is precisely related to the size of the constriction, and in particular to its minimal constriction area. Specifically, if the constriction area is too small, the vocal tract will behave as in the case of plosives; on the other hand, if the constriction area is too wide, no frication noise will be generated and thus an approximant, or frictionless continuant, will be produced.

It appears thus clearly that articulatory control requires a much higher degree of precision for the production of fricatives than for the production of plosives. This fact constitutes indeed an important argument for the explanation of the preponderance of plosives over fricatives not only during the first months of babbling but also in the sound structures of the world's languages.

3.3.2 Voiceless vs. voiced fricatives' predominance

Most fricatives and plosives in the world's languages can be voiced or voiceless. Before discussing issues about voicing, let us state again the basic principles underlying the generation of both voice and frication acoustic source for fricatives. The presence of voicing, i.e. of vocal fold vibration, is mainly controlled by an average pressure drop across the glottis and by vocal fold adduction (or glottis area at rest) (*cf. e.g.* Pelorson et al., 1994). In turn, the pressure drop across the glottis depends on both the glottis area and the geometry of the entire vocal tract. Note in passing that the control of the frequency of vocal folds oscillations is more specifically, but not only, associated with vocal fold length. Finally, the generation of frication is essentially governed by the oral constriction area and by the pressure drop across this constriction (Badin et al., 1995; Stevens, 1998).

The coordination between glottis and oral constriction gestures thus plays a crucial role for acoustic excitation sources in the vocal tract, as has been widely recognised and discussed in the literature (e.g. Scully, 1971; McGowan et al., 1995; Badin et al., 1996; Stevens, 1998). From an aerodynamic perspective, the vocal tract can be viewed as two lumped constrictions between which the subglottal pressure is distributed (*cf. e.g.* Badin et al., 1996; Mawass et al., 2000). Although some refinements could be brought to this crude approximation (*cf.* Pelorson et al., 1994), it is expected that the discussion below will remain valid.

Denoting by A_g the low frequency component of the glottis area, and by A_c the oral tract minimum constriction area, the pressure drops at the glottis ΔP_g and at the constriction ΔP_c are given by the following equations:

$$P_s = \Delta P_g + \Delta P_c, \text{ with } \Delta P_c = \frac{\rho}{2} \cdot \frac{U^2}{A_c^2} \text{ and } \Delta P_g = \frac{\rho}{2} \cdot \frac{U^2}{A_g^2}$$

where P_s is the subglottal pressure, ρ the air density, and U the constant volume flow velocity. It is also known that the amplitude of voicing increases with ΔP_g and reaches a maximum for a given A_g depending on ΔP_g (Stevens, 1998), while the amplitude of the frication noise source is proportional to $\Delta P_c{}^p.A_c{}^q$ (Badin et al., 1995).

It follows that, for a given subglottal pressure, simultaneously increasing voicing and frication amplitudes is contradictory: a balance must necessarily be found for voiced fricative consonants. Preliminary perceptual tests, where the ratio between voicing and frication noise amplitude was varied, showed that the consonants in synthetic vowel-fricative-vowel sequences were deemed acceptable voiced fricatives when the ratio was less than 12 dB (Mawass, 1997). Therefore, simulations performed with a complete vocal tract model, including a simplified aerodynamic model (Badin et al., 1996), were used to determine, for a simplified [ʃ] articulation, the region of the [A_g/A_c] control space where balance between voice and frication noise levels is reached within 12 dB (Mawass, 1997; Abry et al., 1998). The boundaries of this region, schematically drawn from the original simulation data, are shown in Figure 11. This region where the combinations of A_g and A_c values results in a balance between the voice and noise components is indeed very narrow: a rather strict coordination between the glottis and the oral constriction is therefore needed to produce acceptable voiced fricatives (cf. also Mawass et al., 2000). As already noticed by Ohala (1983) on qualitative grounds, this observation contributes to the explanation of the lower proportion of the voiced fricatives (Figure 10) in comparison with the voiceless ones in the world’s languages (Stefanuto & Vallée, 1999).

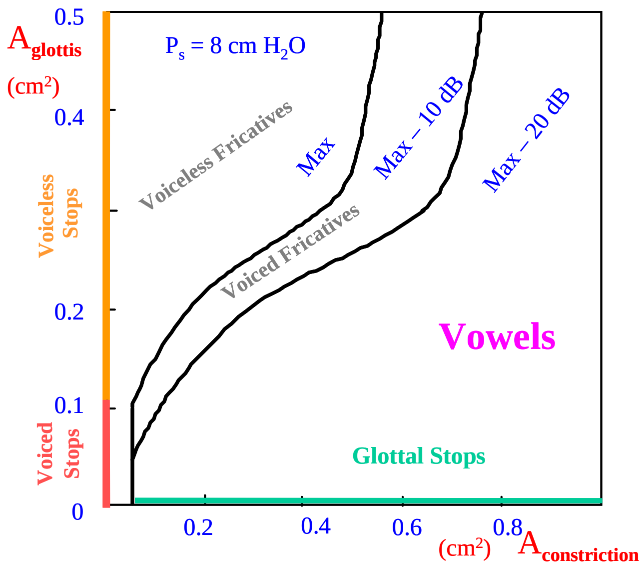


Figure 11: General phase space for manners.
It shows a narrow region in the (A_g/A_c) control space which ensures an approximate balance between voicing and frication noise for fricative consonants
(simplified and adapted from real simulation data, Mawass, 1997; Abry et al., 1998).

4 Discussion and perspectives

A whole collection of data gathered during the course of the second half of the 20th century has progressively rendered untenable the principle of strict independence of form and substance. Languages do not construct their sound systems from “amorphous” materials, to quote a Saussurean expression. Entirely on the contrary, typological analyses carried out on phonological data that have been collected, standardised, and organised, indicate according to current evidence that languages, from whatever linguistic family, do not exploit the possibilities of the vocal tract, auditory and visual systems arbitrarily to organise their phonological structures (be it for the sign of a feature, the choice of that feature, the recruitment of phonemes, or the nature and organisation of syllables). Models that take production and perception constraints into account allow the prediction of broad tendencies and variants for vowel systems. The correspondence between the consonant phonemes that are most frequent in the languages of the world and the consonants that are most often produced during the first stage of babbling, whatever the language of the infant’s native environment, leads to propose the hypothesis that the consonant sounds of different languages are undoubtedly drawn from the stock of potential babbling capabilities.

The phonological typologies and universal tendencies that have been observed are unrelated to the linguistic families which served as the basis for the UPSID samples, currently the most representative database of sound structures. One of the interesting inferences that can be drawn from this lack of correlation is that the general tendencies of linguistic systems could depend on the phonetic substance of the sound inventories; it seems that it might be possible to find confirmation of this in ontogenesis, reinforcing the hypothesis that there might exist “phonetic properties of Homo Sapiens”. Of course, certain phonemes or features (rounding, length, nasality) are associated with certain families or rather with certain geographical areas (for instance /y/ in Indo-European and Uralo-Altaic families). This can only confirm that typological classifications of sound structures and genetic classifications according to linguistic family are far from being identical, and that typology exhibits geographical tendencies. It is likely that statistical properties over the course of time, within a predetermined zone of the linguistic space, reflect the same constraints as the current general distribution, more or less, and that, as regards certain features, fluctuations of systems in neighbouring zones are produced more or less in phase.

Substance-based linguistics is an old dream for phoneticians, expressed in the most vigorous and brilliant terms more than twenty years ago by Bjorn Lindblom, until his recent formula: “derive language from non-language.” Vowel system prediction undoubtedly provides the most obvious success in this program. It is now possible to use a theory based on local and global perceptual (namely “non-linguistic”) arguments, a computational framework, and a quantitative methodology.

The Dispersion-Focalisation Theory (DFT) allows the prediction of vowel systems thanks to competition between two perceptual costs; the associated “phase spaces” determine the DFT winner in this space. We derive from the comparison between experimental phase spaces and UPSID data a region for which the theory predictions fit quite well with the phonological inventories. We have proposed an additional ingredient in our modelling phase-space framework, with the notion of polymorphism of a given phase, or superstructures in phonological organisations. Thanks to this new ingredient, we were able to obtain important results. A next step should consist in adding 8- and 9-vowels systems to the present analysis. Indeed, our analysis of raw $UPSID_{451}$ data demonstrate that more than 9 languages out of 10 (91.5%) in the database contain 3 to 9 vowel qualities. The difficulty however is that phase spaces become increasingly complex with such a large number of vowels.

But whatever the feasibility of this next step, we believe that the present results provide a new illustration of the convincing ability of substance-based theories, from Stevens' Quantal Theory and Lindblom's Dispersion Theory to ICP Dispersion-Focalisation Theory, to produce realistic predictions and useful typologies, at least for vowel systems.

The typology of consonant systems also shows that languages only use a small inventory of consonants even though they potentially dispose of a considerable number of possibilities. The place systems “preferred” by each manner show that languages’ choices are at least partly conditioned by the morphological constraints that facilitate or complicate, make possible or impossible, the articulatory gestures according to the places, e.g. the complexity of the articulatory control of voiced and aerodynamics of fricatives. Although pleading the cause of a non-substantialist linguistics, Trubetzkoy conceded in 1939 (1970 p. 135): *“This fact cannot be ascribed to chance, and must be deeply rooted in the nature of the three series in question.”*

In fact, it is quite likely in the context of the syllable, associating consonants and vowels, that one must elaborate and test the models predicting general trends of sound structures. MacNeilage’s theory, that he describes as a basic principle influencing the very structuring of speech itself, separates the *frame* level (syllabic) and the *content* one (segmental). The nature of mandibular *pure frames* in the canonical babbling stage is placed by MacNeilage in a labial frame. This hypothesis has been tested with 3 individual articulatory models (Vilain, 2000). We think it should rather be placed in a mandibular coronal frame coupled with a visual bilabial frame – which could explain the [d]/[b] coemergence in babbling. This is supported by arguments from articulatory simulations and by our typological arguments on the world’s languages.

Almost thirthy years after Lindblom's and Liljencrant's proposition, the viability of the project of a substance-oriented linguistics seems to be confirmed. Speech is by nature an interdisciplinary object of research, lying at the crossroads of several sensorimotor systems involved in the production and perception of biological communication signals, and of a major human competence, the faculty of language. Around this particular object of research have arisen groupings of several disciplines (ranging from acoustics, even fluid mechanics, to phonology and phonetics, while traversing biology, psychology, and information processing), placing speech at the centre of research that is as much fundamental as it is applied. All of the mutations that phonetics has undergone have provoked a reorganisation of the scientific connections between the phonetic sciences and adjacent disciplines. New interactions are clearly underway between linguistic science, cognitive science, and certain sectors of the physical and engineering sciences.

Hence the importance of "laboratory phonology" which tends to combine experimental phonetics, experimental psychology, and phonological theory (Ohala & Jaeger, 1986). This approach aims to subject hypotheses of phonological organisation to the kinds of validations used in the experimental sciences, which has been lacking until to date in generative phonology. At the end of this century, certain phonologists are thus clearly revising their positions: "*Beyond the differences which traverse it, the current phonological movement tends to give an objective status to the entities which are isolated by analysis. The question of substance, which has for so long been taken to be a secondary issue, only just likely to arouse the interest of certain empirical spirits (phoneticians), is the subject of a renewal of interest. […] The anchoring of the reality of phonological systems in substance now appears to be an indispensable means for thinking of phonology as an interface between physical objects and phonological entities.*" (*De natura sonorum*, Laks & Plénat, 1993). Our approach at the core of the relationships between phonetics and phonology shows that the perception-action interplay contribute to organise the sound structures of the world's languages. Phonetic knowledge can thus explore and make precise the natural constraints that all phonological theories must respect in order to satisfy concerns of (neuro)physiological plausibility. The reintegration of phonology into the natural order of things needs no longer involve a subordinate relationship between the two disciplines.

References

Abry, C., Badin, P., Mawass, K. & Pelorson, X. (1998) The Equilibrium Point Hypothesis and control spaced for relaxation movements or “When is movement actually needed to control movement ?”. *Bulletin de la Communication Parlée*, *4*, 27-33.

Abry, C., Badin, P., Vilain, A., Stefanuto, M. & Boë, L.-J. (1997) Une théorie de l'ontogenèse à l'épreuve de 3 modèles articulatoires anthropomorphiques. In *Un Générateur des Structures Sonores du Langage inspiré des contraintes développementales*. Rapport Final ARASSH (Agence Rhône-Alpes pour les Sciences Sociales et Humaines). Institut de la Communication Parlée, 1998.

Alcantara, V. (1998) *Le trait de labialité dans les langues du monde. De la typologie des systèmes vocaliques à l'enseignement du FLE*,. Travail d'Étude et de Recherche de Maîtrise, Université Stendhal, Grenoble.

Badin, P., Mawass, K., Bailly, G., Vescovi, C., Beautemps, D., & Pelorson, X. (1996) Articulatory synthesis of fricative consonants: data and models. In *Proceedings of the 4th Speech Production Seminar – 1st ESCA Tutorial and Research Workshop on Speech Production Modeling: from Control Strategies to Acoustics*, Autrans, France, 221-224.

Badin, P., Mawass, K. & Castelli, E. (1995) A model of frication noise source based on data from fricative consonants in vowel context. In Elenius, K. & Branderud, P. (Eds.), *13th International Congress of Phonetic Sciences ICPhS95*, Stockholm, Sweden: Arne Strömbergs Grafiska Press, 202-205.

Bailly, G. (1997) Learning to speak. Sensori-motor control of speech movements. *Speech Communication*, 22 (2-3), 251-267.

Boysson-Bardies de, B. (1996) *Comment la parole vient aux enfants*. Odile Jacob, Paris.

Chomsky, N. (1965) *Aspects of the theory of syntax*. MIT press: Cambridge, Mass., USA. French translation: *Aspects de la théorie syntaxique*, Paris: Seuil, 1971.

Chomsky, N. & Halle, M. (1968) *The sound pattern of english*. New York: Harper & Row. *Principes de phonologie générative*. French translation of parts I and IV by P. Encrevé, Paris: Seuil, 1973.

Creissels, D. (1994) *Aperçus sur les structures phonologiques des langues négro-africaines*, 2nd ed., ELLUG, Grenoble, 322 p.

Crothers, J. (1978) Typology and universals of vowel systems in phonology. In *Universals of Human Language* (2), Ed. by H.J. Greenberg, Stanford University Press, Stanford, California, 95-152.

Crystal, D. (1997) *Dictionnary of Linguistics and Phonetics*. Blackwell, Oxford.

Davis, B.L. & MacNeilage, P.F. (1994) Oganization of babbling: A case study. *Language and Speech*, 37 (4), 341-355.

Ducrot, O. & Schaeffer, J.-M. (1995) Nouveau dictionnaire encyclopédique des sciences du langage. Paris: Seuil.

Greenberg J., Ferguson C. & Moravcsik, E. (Ed.) (1978) Universals of Human Languages: Method and Theory, Phonology, Word Structure, Syntax, Stanford University Press.

Hagège, C. (1982) *La structure des langues*. PUF, Paris.

Hockett, C.F. (1955) *A Manual of Phonology*. Waverly Press, Baltimore.

Jakobson, R., Fant, C.G.M. & Halle, M. (1952) *Preliminaries to Speech Analysis*. The MIT Press, Cambridge.

Keating, P.A. (1990) Coronal places of articulation. *UCLA WPP*, 74, 35-60.

Ladefoged, P. & Maddieson, I. (1996) *The sounds of world languages*. Blackwell Publishers, Oxford.

Laks, B. & Plénat, M. (1993) *De natura sonorum*. Presses Universitaires de Vincennes, Vincennes.

Laver, J. (1994) *Principles of Phonetics.* Cambridge University Press, Cambridge.

Liljencrants, J. & Lindblom, B.E.F. (1972) Numerical simulation of vowel quality systems: the role of perceptual contrast. *Language*, 48, 839-862.

Lindblom, B. (1986) Phonetic universals in vowel systems. In *Experimental phonology*, J.J. Ohala (Ed.), New-York: Academic Press, 13-44.

Lindblom, B. & Maddieson, I. (1988) Phonetic universals in consonant systems. In *Language, Speech and Mind. Studies in Honour of Victoria A. Fromkin*. L.M. Hyman & C.N. Li. (Eds.) Routledge, London and New-York, 62-78.

Locke, J.L. (1983) *Phonological acquisition and change*. Academic Press, New-York.

Locke, J.L. & Pearson, D.M. (1992) Vocal learning and the emergence of phonological capacity. A neurobiological approach. In *Phonological development. Models, Research, Implications*, C. Ferguson, L. Menn & C. Stoel-Gammon (Eds.), Timomium, Maryland, 91-129.

MacNeilage, P.F. (1998) The Frame/Content theory of evolution of speech production. *Behavioral and Brain Sciences*, 21 (4), 499-511.

MacNeilage, P.F. & Davis, B.L. (1990) Acquisition of speech production: Achievement of segmental independence. In *Speech Production and Speech Modeling*, W.I. Hardcastle & A. Marchal (Eds.), Kluwer, Dordrecht.

Maddieson, I. (1984) *Patterns of sounds*. Cambridge studies in speech science and communication, Cambridge: Cambridge University Press, 2nd ed. 1986.

Maddieson, I. & Precoda, K. (1989) Updating UPSID, *UCLA WPP*, 74, 104-111.

Mawass, K. (1997) *Synthèse articulatoire des consonnes fricatives du français.* Thèse de doctorat, INP Grenoble.

Mawass, K., Badin, P. & Bailly, G. (2000) Synthesis of French fricatives by audio-video to articulatory inversion. *Acta Acustica*, 86, 136-146.
McGowan, R.S., Koenig, L.L. & Löfqvist, A. (1995) Vocal tract aerodynamics in [aCa] utterances: Simulations. *Speech Communication*, *16*, 67-88.
Ohala, J. (1983) The origin of sound patterns in vocal tract constraints. In *The Production of Speech*, by P.F. MacNeilage (Ed.), Springer Verlag, New-York, 189-216.
Ohala, J. & Jaeger, J.J. (Eds.) (1986) *Experimental phonology*. Academic Press Inc., Orlando.
Pelorson, X. (1997) On the fluid mechanics of bilabial plosives. *Speech Communication*, *22*, 155-172.
Pelorson, X., Hirschberg, A., Van Hassel, R.R., Wijnands, A.P.J. & Auregan, Y. (1994) Theoretical and experimental study of quasi-steady flow separation within the glottis during phonation. Application to a modified two-mass model. *Journal of the Acoustical Society of America*, *96*, 3416-3431.
Prince, A. & Smolensky, P. (1993) *Optimality theory: constraint interaction in generative grammars.* ms, Rutgers University, Université du Colorado, New Brunswick et Boulder.
Robb, M. & Bleile, K. (1994) Consonant inventories of young children from 8 to 25 months. *Clinical Linguistics and Phonetics*, 8, 295-320.
Saussure, F. De (1916) *Cours de linguistique générale*. Publication de Ch. Bally & A. Sechehaye, Payot, Lausanne, Paris. Reedition by De Mauro, T., Payot, Paris, 1972).
Scully, C. (1971) A comparison of [s] and [z] for an english speaker. *Language and Speech*, 14, 187-200.
Schwartz, J.L., Abry, C., Boë, L.J., & Cathiard, M. (2002) Phonology in a theory of perception-for-action-control. In *Phonology : from Phonetics to Cognition*, Durand J. & B. Laks (Eds.), Oxford University Press, Oxford, 255-280.
Schwartz, J.-L., Boë, L.-J., Vallée, N. & Abry, C. (1997a) Major trends in vowel system inventories. *J. of Phonetics*, 25 (3), 233-253.
Schwartz, J.-L., Boë, L.-J., Vallée, N. & Abry, C. (1997b) The Dispersion-Focalization Theory of vowel systems. *J. of Phonetics 25* (3), 255-286.
Schwartz J.L., & Escudier, P. (1987). Does the human auditory system include large scale spectral integration?. In *The Psychophysics of Speech Perception*, M.E.H. Schouten (Ed.), Nato Asi Series, Martinus Nijhoff Publishers, Dordrecht, 284-292.
Schwartz, J.L., & Escudier, P. (1989). A strong evidence for the existence of a large scale integrated spectral representation in vowel perception. *Speech Communication*, 8, 235-259.
Sedlak, P. (1969) Typological considerations of vowel quality systems. *Working Papers on Language Universals*, 1, Stanford University, 1-40.
Shadle, C. (1990) Articulatory-acoustic relationships in fricative consonants. In *Speech Production and Speech Modelling*, W. Hardcastle and A. Marchal (Eds.), Kluwer Academic Publisher, 187-209.
Stefanuto, M. (1996) *Typologie des lieux d'articulation des langues du monde.* Travail d'Étude et de Recherche en Sciences du Langage, Université Stendhal, Grenoble.
Stefanuto, M. (1997) *Apport de la neurocognition comparée, développementale et pathologique dans l'étude des tendances universelles phonologiques*. Mémoire de DEA en Sciences du Langage, Université Stendhal, Grenoble.
Stefanuto, M. & Vallée, N. (1999) Consonant systems: From universal trends to ontogenesis. *Int. Congr. of Phonetic Sciences ICPhS99* (3), San Francisco, 1973-1976.
Stevens, K.N. (1972) The quantal nature of speech: Evidence from articulatory-acoustic data. In *Human communication: A unified view*, P.B. Denes & J.R. Davis (Eds.), McGraw-Hill, New York, 51-66.
Stevens, K.N. (1998) *Acoustic Phonetics*. The MIT Press: Cambridge, MA, London, England.
Stoel-Gammon, C. (1985) Phonetic Inventories 15-24 months: A longitudinal study. *J. of Speech and Hearing Research*, 23, 506-512.
Trubetzkoy, N.S. (1939) Grundzüge der Phonologie. *Travaux du Cercle Linguistique de Prague*, 7, 272 p. Translation by Cantineau, J. (1970) *Principes de phonologie.* Klincksieck, Paris, 394 p.
Vallée, N. (1994) *Systèmes vocaliques: de la typologie aux prédictions.* Thèse de Doctorat en Sciences du Langage, Université Stendhal, Grenoble.
Vallée, N., Schwartz, J.-L. & Escudier, P. (1999) Phase spaces of vowel systems. A typology in the light of the Dispersion-Focalization Theory (DFT*). Int. Congr. of Phonetic Sciences ICPhS99* (1), San Francisco, 333-336.
Vihman, M.M. (1996) Phonological development: The origin of language in the child. Blackwell Publishers.
Vilain, A. (2000) Apports de la modélisation des degrés de liberté articulatoires à l'étude de la coarticulation et du développement de la parole. Thèse de doctorat, Université Stendhal, Grenoble.
Vilain, A., Abry, C., Brosda, S. & Badin, P. (1999) From idiosyncratic pure frames to variegated babbling: Evidence from articulatory modelling. *Int. Congr. of Phonetic Sciences ICPhS99* (4), San Francisco, 2497-3500.