The Role of Release Bursts in Final Stop Perception*

Paul Marty
Massachusetts Institute of Technology

ABSTRACT

The absence of release bursts in final stops has often been proposed as the factor responsible for the neutralization of place contrasts. Specifically, in languages where final stops are *never* released (e.g., Cantonese), it has been observed that place contrasts neutralize when formant patterns are acoustically too similar (e.g., up˺-uk˺). This paper focuses on the role of release bursts in phonological structure and speech perception by investigating how native speakers of French, a language where final stops are *always* released, perceive bursteless stops. Results support the view that release bursts are not only useful, but critical to place distinctions in environments where formant transitions alone would be insufficiently distinct. Results also reveal that, when perceptual confusion occurs, French listeners default neither to the linguistically unmarked value, nor to the most frequent value in context. Rather, they overwhelmingly default to the place feature for which the release burst is most similar acoustically to no burst at all. Within a production-perception grammar, I propose an account in terms of cue weighting, modulated by listeners' expectations about the presence of the relevant place cues.

1 Introduction

This paper proposes to consider from a new angle an old issue in speech perception, namely the contribution of release bursts to the perception of post-vocalic final stops. Release bursts are known to play an important role in the discrimination and identification of stop consonants (Householder 1956, Dorman *et al.* 1977, Ohde & Stevens 1983, Kewley-Port *et al.* 1983). It has been observed for languages where coda stops are *never* released (e.g., Cantonese, Taiwanese, Hakka) that place contrasts neutralize in environments where formant transitions would otherwise be insufficiently distinct (Kawasaki-Fukumori 1992, Flemming 2002, Flemming 2004). Based on these observations, the absence of release bursts in the production of final stops was proposed as a critical factor in the neutralization of place contrasts (Ohala 1993, Ohala 2003, Jun 2004). However, little experimental evidence has been offered thus far in support of the proposal that release bursts are not only useful, but essential to place distinctions in those environments. The present study intends to provide a direct test of this claim. Before going on, let us succinctly describe the distribution of the two major sources of cues to place of articulation in VC sequences.

*I am indebted to Edward Flemming, Peter Graff, Michael Kenstowicz and Donca Steriade for suggestions and comments on earlier versions of this work. I would also like to thank audience at CLS 48 for useful feedback. All errors are my own.

First, the place of C is cued by formant transitions. In VC sequences, articulators move from a vocalic to a consonantal position, modifying the shape of the vocal tract. These changes are acoustically reflected in the transition of the second formant (F2) - and of the third formant (F3) to a lesser degree - during vowel offset (Fant 1970), and provide cues to the identity of the adjacent consonantal constriction (cf. Delattre *et al.* 1955, Liberman *et al.* 1958). Secondly, the place of C may have internal cues, namely the frequency composition and amplitude of C's burst (Blumstein *et al.* 1979). The availability of this source of information varies across-languages. For instance, final stops are always released in French or Russian, variably released in English, and never released in Cantonese or Taiwanese Mandarin. Previous studies in speech perception have shown that there are robust phonetic differences between released stops and unreleased stops, the former being more perceptually salient and less confusable than the latter (Malécot 1958, Winitz *et al.* 1972, Wright 2004, a.o.). Results from Kochetov & So (2007) suggest further that the information provided by burst cues is beneficial both to listeners that are accustomed to final stop releases in their native language and to those who are not.

The hypothesis that the lack of burst cues drives neutralization of place contrasts captures the trend towards neutralizing place contrasts in languages where coda stops are *never* released. To illustrate this neutralization pattern, consider the frequent co-occurrence restrictions on coda labials (or labialized consonants) in rounded vowel environments. Cantonese is a language that exhibits such restrictions. While back round vowels can appear with a labial onset, they do not appear with a labial coda (cf. Cheng *et al.* 1991):[1]

(1)	CV	VC
	pu	*up˺
	ku	uk˺
	tu	ut˺

The diachronic study of sound change in Cantonese reveals that these restrictions result from the neutralization of consonant contrasts. Back rounded vowels do not appear preceding labials because labials have been replaced by velars (i.e., up˺> uk˺). Parallel to this is the observation that the formant transitions leading into labials and velars are acoustically very similar when adjacent to back rounded vowels (Öhman 1966, Olive *et al.* 1993) - too similar arguably to support the /p-k/ contrast in the absence of audible releases. As Flemming (2002) observed, the availability of burst cues in initial stops can account for the directionality of these co-occurrence restrictions: while the relevant C place contrast is neutralized in VC contexts, it is maintained in CV contexts by the differences in the quality of the release bursts.

This pattern of neutralization is not limited to Cantonese. In a large number of Chinese languages, all stop and nasal place contrasts are neutralized in coda position (cf. Cheng 1973, see also Chen & Wang 1975), and the result is always a velar (e.g., Hakka) or a glottal (e.g., Highland Yao). To a large extent, neutralization

[1]Coda stops in Cantonese are glottalized and thus generally unreleased. Following the IPA, we will indicate the lack of an audible release burst using an upper-right corner diacritic (˺) above the relevant stops.

can be thought of as revealing invariants in preferences and dis-preferences among sound sequences. Moreover, the direction of neutralization can be identified by reference to phonetic implementation factors. According to Kawasaki-Fukumori (1992), these preferences (and dis-preferences) can be accounted for by means of the following two acoustic factors: i) the degree of acoustic difference between sequences, and ii) the magnitude of acoustic modulation within a sequence. Let us rely on these factors to capture within a perceptual grammar the propensity to neutralize velars in back-rounded environments.

The first factor relates directly to the requirement that the distinctiveness of contrasts be maximized. For the present purposes, this requirement can be implemented using a MINDIST constraint that penalizes place contrasts that are not distinct enough regarding the combination of formant and burst differences (cf. Flemming 2002).[2] The second factor is relevant given the hypothesis that modulation along acoustic dimension (e.g., frequency, amplitude, etc.) should result in increased salience of the cues in the portion of the signal where the change occurs. Following this hypothesis, we will assume that the lack of modulation between a rounded vowel and a labial stop makes the /up˺/ sequence less different from the /u/ than the /uk˺/ sequence is from the /u/ - implying thus that language would exert less effort in maintaining a labial stop in this context (as compared to a velar stop). Such a preference can be formalized using the *ad hoc* constraint ranking *p >> *k. Finally, we will use MAXIMIZE CONTRAST to refer to the preference for more contrasts (cf. Flemming 2002). The tableau in (2) illustrates the crucial rankings between the relevant constraints.

(2)

			MINDIST F2-F3 or burst diff	MAXIMIZE CONTRASTS	*p	*k
a.		u-up˺-ut˺-uk˺	*!	✓✓✓✓	*	*
b.		u-up˺-ut˺		✓✓✓	*!	
c.	☞	u-ut˺-uk˺		✓✓✓		*

Place neutralization occurs because the contrast between labials and velars in the [u] environment cannot be supported in the absence burst cues. The a-candidate is eliminated by one violation of the MINDIST constraint, due to the acoustic closeness of formant patterns between up˺and uk˺. While the remaining b- and c-candidates equally satisfy MAXIMIZE CONTRASTS, the b-candidate is ruled out by the ranking *p >> *k, which is assumed to reflect the preference for sound sequences exhibiting greater perceptual salience relative to /u/. The c-candidate ends up being the winning output, as expected. Hence, this phonetic account accurately predicts both the neutralization of the labial-velar contrast in the [u] environment and the output of this phonological process in languages such as Cantonese (i.e., labials are replaced by velars and not the other way around). It further accounts for the preservation of this contrast in languages such as French where the availability of release

[2]For the sake of simplicity, we present here a non-quantitative version of the MINDIST analysis that takes distances in F2-F3 transitions falling below some threshold to be distances of 0. See Flemming (2001) for a different implementation of this idea using weighted constraints.

bursts in final stops contributes to place distinctions, as illustrated in (3).

(3)

			MinDist F2-F3 or burst diff	Maximize contrasts	*p	*k
a.	☞	u-up-ut-uk		✓✓✓✓	*	*
b.		u-up-ut		✓✓✓!	*	
c.		u-ut-uk		✓✓✓!		*

Patterns of place neutralization such as the one described above have been taken to indicate a correlation between the absence of release bursts and stop contrast neutralization. The present paper offers a direct test of the hypothesis that place neutralization in word-final contexts is driven by the absence of release bursts. This is done by testing how native speakers of French, where final stops are obligatory released, perceive final stops when we take out burst cues from the acoustic signal. The prediction is the following: in the absence of burst cues, French listeners should perceptually neutralize place contrasts that are acoustically too close in terms of F2-F3 transitions. The paper is organized as follows. Section 2 reports on an acoustic study showing that some formant patterns found in languages that exhibit categorical neutralization of place contrasts are also observed in French. Section 3 reports on a phoneme-identification experiment which tests the extent to which the removal of release bursts affects French listeners' perception of the relevant contrasts. Overall, acoustic and perceptual data provide evidence that neutralization and the absence of burst cues can be linked in the light of formant similarities. The last section proposes a preliminary account for the particular direction of the perceptual confusion observed in French listeners.

2 Acoustic Study

This study examines the character of the VC transitions leading into the three places of articulation in French (i.e., labial, coronal and velar) and their associated release bursts. It aims at identifying specific environments in which perceptual neutralization would be likely to occur in the absence of burst cues.

2.1 Materials

9 nonsense VC syllables (C=/ptk/, V=/aiu/) were recorded in the sound attenuated booth of the MIT Phonetics Laboratory. The microphone was placed about 15 cm from the talkers' mouth and slightly off the right side so as to avoid artifact. Talkers were 2 native speakers of French (2 males; mean age 25 years), who were born and raised in France. They were recorded uttering the target VC sequences in random order in the carrier sentence [Regarde, il y a le __] ('Look, there is the __'). Each utterance was repeated 6 times.

2.2 Method

Measurements of the formant transitions and release bursts were performed on the recordings of the two speakers, including thus 108 tokens (9 sequences $\times$ 6 repetitions $\times$ 2 speakers). Linear prediction (LPC) analysis produced by the PRAAT software (Boersma & Weenink 2008) was used to provide detailed temporal and spectral measurements of the formant transitions (sample window of 25 ms). The acoustic properties of release bursts were determined based on the study of their waveforms.

2.3 Predictions

The methodological choice of the three vowel contexts /a, i, u/ capitalized on previous observations about the presence or absence of similarities between formant transitions in these environments (Halle *et al.* 1957, Delattre 1958, Öhman 1966, Olive *et al.* 1993, Ohala & Ohala 1998). Specifically, VC transitions have been found to be most different in the environment of the open vowel [a], while they cross-linguistically tend to show similarities in the environment of the high vowel [i] (for labials and coronals) and of the back rounded vowel [u] (for labials and velars). Relying on these generalizations, we expected to observe acoustic closeness between distinct places of articulation in these specific vowel contexts. This result would provide us with cases of primary interest to test the hypothesis that, in the absence of release bursts, acoustic similarities in VC transitions drive perceptual confusion of place contrasts.

2.4 Results

2.4.1 Formant Patterns

VC transitions were extracted by the PRAAT LPC-based formant algorithm. The acoustic analysis aimed at determining quantitatively the changes occurring during the period of co-articulation between the vowel and the consonantal stop by measuring F2 and F3 frequencies at the midpoint and endpoint of the relevant vowels. We report in Figure 1 the frequency differences in absolute values for F2 and F3 at vowel endpoint between the three pairs of sounds as a function of vowel contexts: the smaller the bar is, the greater the acoustic closeness.

Overall, formant patterns conform to expectations, based on patterns seen in other languages (cf. Halle *et al.* 1957, Delattre 1958, Fant 1970, Olive *et al.* 1993, Ohala & Ohala 1998). In the [a] environment, places of articulation were found to be well differentiated by distinct frequencies of F2 and F3 at vowel midpoint (all $|\Delta F2|$ and $|\Delta F3| \geq 190$ Hz). By contrast, the analysis revealed noticeable similarities between patterns of the transitions in [i] and [u] vowel contexts. First, labials and coronals exhibited very similar F2 and F3 patterns after [i] ($|\Delta F2|$ and $|\Delta F3| \leq 80$ Hz), whereas VC transitions to these two places of articulation were distinct in the other environments. Second, F2 and F3 frequencies for labials and coronals were very close after [u] ($|\Delta F2|$ and $|\Delta F3| \leq 100$ Hz): transitions for these two sequences were essentially flat and only differ in that F2 values were slightly higher at vowel endpoint for [uk] than for [up]. These findings confirm that VC transitions leading into the three places of articulation are distinct after [a], and reveal acoustic closeness between labials an coronals after [i] and between labials and velars after [u].

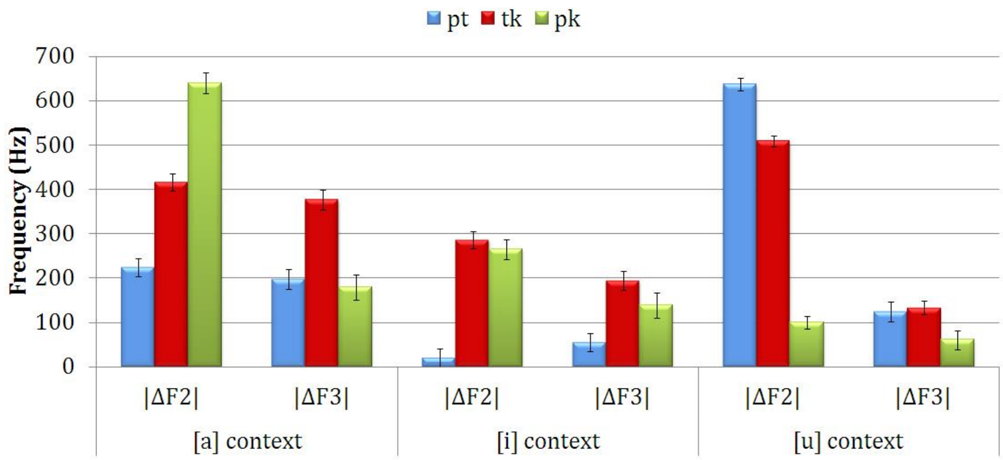

Figure 1: Frequency differences (Hz) in absolute values for F2 and F3 at vowel endpoint between the three pairs of sounds (pt, tk, pk) as a function of vowel contexts (a, i, u). Error bars refer to standard errors.

2.4.2 Acoustic properties of release bursts

All the recorded VC tokens showed audible C release bursts. Spectra observed for consonantal releases revealed a distinctive shape for each place of articulation: a diffuse spectrum for labials, a diffuse-rising spectrum for coronals and a prominent mid-frequency spectral peak for velars (cf. Blumstein *et al.* 1979). Velar and coronal bursts were found to be longer in duration than labial bursts (on average 11 ms and 12 ms for [k] and [t] *vs* 7 ms for [p]). Furthermore, velar bursts were also higher in amplitude than coronal and labial bursts (about 3 dB and 6 dB higher, respectively). Taken together, these data allow us to establish that velar bursts are the loudest and longest, followed by coronal and then labial bursts. This acoustic scale is consistent with previous cross-linguistic observations and confirms that labials have a diffuse, low intensity burst, while back-articulated velars tend to have a compact, relatively intense burst by virtue of their downstream resonators.

2.5 Summary

The present results make two contributions. First, they show that velar bursts are acoustically the most prominent, followed by coronal and then labial bursts. Second, they show that the acoustic distance between the formant transitions leading into the three places of articulation in French substantially varies as a function of the vowel environment: while the formant transitions leading into /p, t, k/ are very different from each after [a], similarities in F2 and F3 patterns for ip˺-it˺and up˺-uk˺indicate that these pairs of sounds are likely to trigger perceptual confusion by virtue of their acoustic closeness. These observations are consistent with formant patterns found in other languages. In the following perceptual experiment, we propose to use these VC sequences to address directly the question of the contribution of release bursts to the preservation of place contrasts for final stops. It should allow us to determine whether the contribution of the bursts is critical for identifying accurately place of articulation in [i] and [u] environments.

3 Perceptual Study

This study examines how well the three voiceless stops in French are perceptually differentiated in VC environments both with and without their release bursts. Following Ohala (1989, 1993, 1998), we expected French listeners to be unable to differentiate burstless stops in environments where VC transitions alone are not distinct enough to allow place contrasts. Results provide support for this prediction. Unexpectedly, in case of perceptual confusion, we found that French listeners overwhelmingly default to the place feature for which the release burst is most acoustically similar to a burstless segment. A post-hoc explanation of these findings is proposed in the last section.

3.1 Materials

Test items were natural speech VC tokens (C=/ptk/, V=/iu/). They were selected among the two male speakers' recordings (cf. previous section) for exhibiting maximally distinct formant patterns for the putatively confusable pairs of sounds [ip]-[it] and [up]-[uk]. In addition to the test items, tokens of [ap], [at] and [ap] were included as fillers to disguise the goal of the experiment. Filler items were not considered for analysis.

3.2 Method

Stimuli were used in two conditions: CONTROL and TARGET. Stimuli for the CONTROL condition were edited from the recorded speech materials. Potential information provided by noise occurring during stop closures (Repp 1983, Chitoran *et al.* 2002) was eliminated from the acoustic signal by setting the amplitude during the closure of C to zero. Stimuli for the TARGET condition were then edited from the CONTROL stimuli by truncating the closure and release of C from the acoustic signal. The editing manipulations were all made at audio waveform zero-crossings. Following this procedure, the place of C was cued both by VC transitions and the release of C in the CONTROL condition, whereas it was only cued by VC transitions in the TARGET condition. To manipulate the perceptual cue factor within subjects, listeners were administered two consecutive blocks of trials: one block contained the CONTROL stimuli, and the other block contained the TARGET stimuli.

3.3 Listeners

The listeners were 12 native speakers of French (6 males), aged between 22 and 29 years, who were naive as to the purpose of the experiment. They participated in this study on a voluntary basis. None of the listeners reported hearing problems.

3.4 Procedure

The experiment was a forced-choice phoneme-identification task. The stimuli were presented individually via headphones at equal volumes to both ears, and at a self-selected comfortable intensity level. The listeners were instructed to listen to speech tokens and to identify the consonant (i.e., 'p', 't' or 'k') by pressing arrow keys (i.e., $\triangleleft$, $\triangledown$ or $\triangleright$) on a standard keyboard. They were asked to keep their fingers on the response keys throughout the experiment, and to respond as quickly and accurately as possible. Response and key arrangements were counter-balanced across participants. Listeners were administered two blocks of 54 trials (9 sequences $\times$ 2 speakers $\times$ 3 repetitions) with a brief self-time break in between. For each listener, it was pseudo-randomly determined which type of stimuli block (CONTROL or TARGET) they started with. In each block, stimuli were presented in random order with an interstimulus interval randomized between 1250 ms and 1750 ms. No feedback was given regarding the responses during the sessions. Each session lasted approximately 10 minutes.

3.5 Prediction

The prediction is that listeners should be unable to identify accurately the consonants in the TARGET condition (i.e., in the absence of burst cues) for the pairs of sounds [up˺]-[uk˺] and [ip˺]-[it˺], i.e. when VC transitions exhibit substantial acoustic similarities. Such results would provide strong support to the view that burst cues are not only useful in these environments, but critical.

3.6 Data Collection and Analysis

A total of 850 responses to test items were collected.[3] Responses are summarized in the three-category confusion matrix given in Table 1. First, listeners' responses in the CONTROL condition were examined to check if participants did the task appropriately. With a global mean score of 98% correct trials, it appears that listeners fully relied on the combination of VC transitions and C bursts to identify the target consonants. Second, listener's responses in the TARGET condition were analyzed to answer the following two questions: (i) Did French listeners hear a sound different from the one they were presented with? (ii) If so, which place value did they default to?

To address both questions, the confusion matrix in Table 1 was analyzed in terms of the biased choice model (Luce 1959, Luce 1963) by fitting a loglinear model to listener's responses for the TARGET condition. Specifically, the bias choice model (BCM) returns a symmetrical confusability estimate for each unordered pair of sounds (perceptual distance), while controling for response biases by estimating a separate model parameter for them (bias parameters).[4] Perceptual distances were used to quantify the confusability generated by similarities in VC transitions in the

---

[3]Due to a technical error, one participant was presented with 2 instead of 3 repetitions of the test items, resulting in a loss of 24 responses.

[4]The bias choice model is given as follows (Luce 1963):

$$P(R_j|S_i) = \frac{\beta_j \eta_{ij}}{\sum_k \beta_k \eta_{ik}}$$

The probability $P(R_j|S_i)$ that a stimulus S_i will incur a response R_j is defined as the response bias β_j times the symmetric similarity of S_i and R_j, η_{ij}, normalized by a constant equal to the sum over all response biases and similarities between S_i and all possible responses. These parameters may be estimated using maximum likelihood fitted log-linear regression predicting the log-number of responses of one sound given another ($\log(R_j|S_i)$).

absence of C bursts, and bias parameters to determine whether listeners default to a particular place value.

Condition	Context (V)	Consonant (C)	Responses 'p'	't'	'k'	% correct
CONTROL	i	p	**64**	3	3	92
		t	1	**69**	0	98
		k	0	0	**70**	100
	u	p	**70**	0	0	100
		t	1	**67**	2	96
		k	1	0	**69**	98
					means	97
TARGET	i	p	**61**	7	2	87
		t	21	**45**	4	64
		k	13	13	**44**	63
	u	p	**65**	1	4	93
		t	3	**65**	2	93
		k	56	4	**10**	14
					means	69

Table 1: Responses for the VC sequences in the CONTROL and TARGET condition as a function of the V context and the following C.

3.7 Results

The Maximum-Likelihood estimates of the BCM distance and bias parameters are provided in Figure 2. For the BCM distances, small numbers correspond to high confusability or low perceptual distinctiveness (e.g., distinctiveness of [up˺]-[uk˺] equals 0.49), while large numbers correspond to low confusability or high perceptual distinctiveness (e.g., distinctiveness of [up˺]-[ut˺] equals 3.77). Model comparisons were performed using the likelihood-ratio statistic to test whether distance parameters for the pairs p-t, p-k and t-k in contexts differ from each other and from zero. In each of the two vowel contexts, distance parameters were significantly different from each other ($p < .0001$). The lowest perceptual distances were observed between labials and coronals after [i], and between labials and velars after [u]. All distance parameters were found to be statistically different from zero ($p < .0001$), except for the pair [up˺]-[uk˺] whose likelihood ratio did *not* reach significance ($\chi^2 = 3.30,\ p = .08$). These results indicate that, in the absence of release bursts, French listeners perceptually merged the labial-velar contrast in the [u] environment.

Bias parameters for the three response categories ('p', 't' and 'k') were first compared to each other on the basis of the Wald statistics (z value) associated to each coefficient. Their orderings were found to be the same in both vowel contexts: $b_p > b_t > b_k$. This finding suggests that listeners overwhelmingly default to 'p'. To test whether the absence of bursts biases listeners towards a 'p' response, incorrect responses were all fit into a linear mixed modal (binomial family) that examined the effects of V context and C place on the mean 'p' responses. None of the main effects reach significance ($\beta = 0.61,\ z = 0.72$ and $\beta = -1.11,\ z = -1.54$, ns), but their interaction did: $\beta = 6.34,\ z = 2.58,\ p < .01$. The direction of this interaction is that listeners provided more 'p' responses in cases of lower perceptual distinctiveness, that is more 'p' responses for [k] than for [t] after [u] (93% *vs* 50%) and for [t] than for [k] after [i] (84% *vs* 60%).

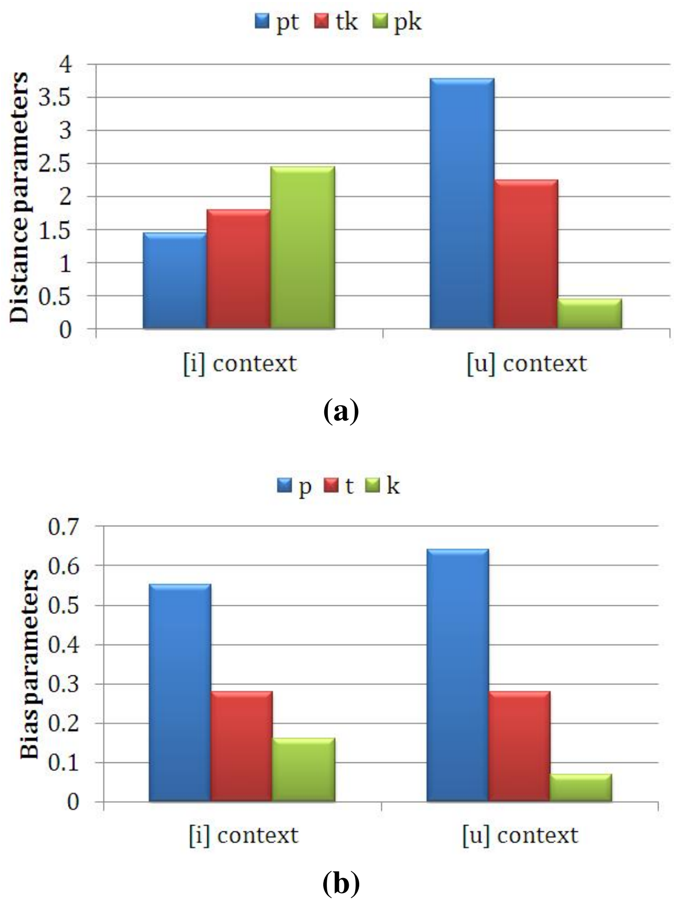

Figure 2: (a) Perceptual distances and (b) Bias parameters.

3.8 Discussion

As noted by Ohala (1993), informed reading of the record of sound change can provide useful hints about speech perception which can then be followed up by perceptual studies in the laboratory. First, the present results clearly show that the labial-velar distinction is eliminated in the [u] context when release bursts have been removed from the acoustic signal. Second, we found that the labial-coronal contrast in the [i] context goes the same way but to a less severe degree. Overall, these findings provide evidence for the view that place neutralization in VC sequences is driven by the absence of burst cues in environments where formant transitions are confusable, and extend this view to French where final stops are always released.

Our results also show that French listeners have a bias in favor of hearing [p] when they are presented with a burstless stop. Precisely, listeners appear to have a pro-p bias as much as they have an anti-t and anti-k bias (in the [u] and [i] contexts, respectively). In the last section of this paper, we discuss concurrent explanations to account for the place value French listeners default to in cases of perceptual confusion.

4 Pos-hoc Investigations

A large number of languages where final stops are unreleased neutralize the /p-k/ contrasts in rounded vowel environments, and the output of this categorical neutralization is often a velar. The goal of this section is to examine possible explanations to account for the different place default in speakers of French. As a first attempt, we checked whether French listeners default to the the most frequent final stops in their own lexicon. For these purposes, we examined the lexical frequencies of our target VC sequences in the French lexicon using the database Lexique 3 (www.lexique.org). As illustrated in Figure 3-(a), results showed that in-context lexical frequencies cannot account for the direction of the perceptual confusion observed in our experiment: while there is a small advantage of /up/ over /uk/ (about 20 tokens), this proposal would not be consistent with the observation that the count of /it/ largely dominates this of /ip/.

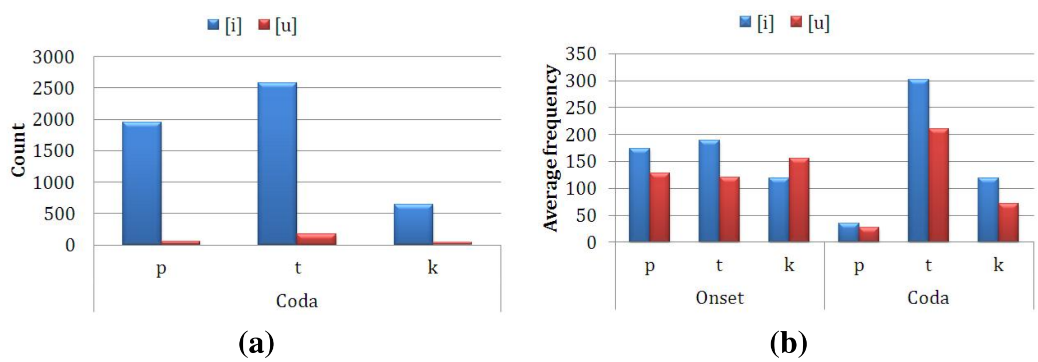

Figure 3: (a) Count (French lexicon) of /p, t, k/ in post-/i, u/ contexts (Coda), and (b) Average frequency (35 languages) of /p, t, k/ in pre- and post-/i, u/ contexts (Onset *vs* Coda)

Secondly, we checked whether our data could reflect different preferences among sound sequences in languages where final stops are released. We investigated this second possibility by carrying out a large scale typology based on the study of 35 languages included in the WOLEX corpus (Graff *et al.* 2011, www.wolex.org), all attested to have final stop releases (either obligatory or optionally). Results reported in Figure 3-(b) showed that cross-linguistic frequencies cannot account for the direction of perceptual confusion in French listeners. In fact, the lexical frequencies observed in languages with final stop releases mirror the propensity for neutralizing coda labials (as compared to velars) that is found in the categorical typology of languages with unreleased final stops (e.g., Cantonese, Hakka): in substance, ‘p’ was found to be much more frequent in onset than in coda position, and largely dis-preferred over ‘k’ (or ‘t’) in coda position. Hence, neither language-specific frequencies, nor cross-linguistic frequencies appear to provide a satisfying explanation for the current data.

Our proposal is to account for the default in French listeners in terms of input-to-output mapping process. As adult listeners, French listeners are optimal listeners: they have learned to minimize the probability of misidentifying an auditory event by classifying it as the category that is most likely to occur given prior knowledge and what is given in/by the acoustic signal (cf. Clayards *et al.* 2008, Boersma & Hamann 2008, Clayards 2010). Assuming that French listeners always expect the presence of a release burst - since final stops are always released in French - we argue that the *absence* of release bursts in the acoustic signal may in fact provide cues, namely for the burst which is acoustically most similar to no burst at all. This conjecture is directly supported by the observation that the acoustic prominence and robustness of the relevant release bursts (i.e., $p < t < k$) is inversely correlated with the ordering of the bias parameters provided by the BCM analysis (i.e., $b_p > b_t > b_k$). This suggests that, in case of perceptual confusion, French listeners perceive a burstless stop as the burst percept which presents the greatest match with the weakest possible burst. Let us finally flesh out this proposal in a more formal way by using a production-perception grammar that maps a fully specified phonetic form onto a fully specified surface form, i.e. the percept, and the following set of constraints:

- *NOBURST: Penalize burstless percepts for final stops. Assign ‘*’ for each final stop represented as burstless in the surface form. This constraint formalizes listeners’ expectations relative to final stop releases in their production grammar.
- MAXDIST(F2:F3): The formant values of the surface form have to be acoustically close to those of the phonetic form. Assign ‘*’ when both $|\Delta$ F2$|$ and $|\Delta$ F3$|$ are superior to 100 Hz.
- DEP(BURST): Do not epenthesize a burst into the surface form. The number of violations is assigned according to the acoustic distance of the relevant burst relative to 0-amplitude: ‘*’ for p-burst, ‘**’ for t-burst and ‘***’ for k-burst.

The tableaux in (4)-(7) illustrate the crucial rankings between the three constraints stated above. Consider first cases that did not give rise to perceptual confusion in our experiment. Consistent with the perceptual data reported above, our grammar predicts in these cases that perceptual confusion should not occur, and that listeners should map burstless phonetic forms onto their corresponding burst surface forms. For instance, we found that 47% of the responses for the stimulus [ik˺] were not-‘k’, with no preference among these for ‘p’ as opposed to ‘t’ (see Table 1). Our proposal is consistent with these findings as illustrated in (4): MAXDIST(F2:F3) rules out both |ip˂| and |it˂| as these percepts are acoustically distinct from [ik˺] in terms of F2-F3 transitions, and [ik˺] is finally matched to |ik˂| at the cost of three violations of the lower-ranked DEP(BURST) constraint. A similar explanation holds for the case of [ut˺], as illustrated in (5).

(4)

[ik˺]			*NOBURST	MAXDIST (F2:F3)	DEP(BURST)
a.		|ik˺|	*!		
b.		|ip˂|		*!	*
c.		|it˂|		*!	**
d.	☞	|ik˂|			***

(5)

[ut˺]			*NOBURST	MAXDIST (F2:F3)	DEP(BURST)
a.		|ut˺|	*!		
b.		|up˂|		*!	*
c.	☞	|ut˂|			**
d.		|uk˂|		*!	***

Consider now the place contrast between [up˺] and [uk˺] which has been found to be perceptually merged by French listeners (see tableaux in (6) and (7) below). In both cases, the most faithful 'burst' percepts, i.e. the a-candidates, are ruled out by the markedness constraint *NOBURST, which is taken to be high-ranked in the production-perception grammar of French listeners.[5] Next, |ut˂| is ruled out by MAXDIST(F2:F3) as it doesn't exhibit formant patterns similar to those observed in the phonetic forms (see Figure 1). Finally, the input-to-output mapping process ends up eliminating |uk˂| and selecting |up˂| due to a difference in the number of violations of DEP(BURST), which reflects the fact that epenthesizing a p-burst (less prominent) is more consistent with the absence of bursts in the acoustic signal than epenthesizing a k-burst (more prominent). Hence the direction of perceptual confusion in French listeners.

(6)

[up˺]			*NOBURST	MAXDIST (F2:F3)	DEP(BURST)
a.		|up˺|	*!		
b.	☞	|up˂|			*
c.		|ut˂|		*!	**
d.		|uk˂|			**!*

[5] Since *NOBURST encodes listeners' expectations relative to release bursts in final stops, the current model predicts that *NOBURST should be low-ranked in languages that lack release bursts for final stops such as Korean (and DEP(BURST) inactive then), and that its ranking should depend on the specific conversational circumstances in languages with optional releases such as English.

(7)

[uk˺]			*NOBURST	MAXDIST (F2:F3)	DEP(BURST)
a.		|uk˺|	*!		
b.	☞	|up˂|			*
c.		|ut˂|		*!	**
d.		|uk˂|			**!*

5 Conclusion

A number of studies in phonology have investigated the effect of the presence or absence of different cues on performance and linguistic typology. The current study presents evidence for the potential role of listeners' expectations interacting with the presence or absence of cues in the acoustic signal. This proposal makes a direct prediction: listeners' expectations can be modulated by modifying the likelihood distribution of burst cues in the acoustic signal. This conclusion is consistent with an increasing body of evidence supporting the role of conditional probability in linguistic behavior. We hope our work will contribute to effort in this direction.

References

Blumstein, S.E., K.N. Stevens, *et al.*. 1979. Acoustic invariance in speech production: Evidence from measurements of the spectral characteristics of stop consonants. *Journal of the Acoustical Society of America* 66.1001–1017.

Boersma, P., & S. Hamann. 2008. The evolution of auditory dispersion in bidirectional constraint grammars. *Phonology* 25.217–270.

——, & D. Weenink. 2008. Praat: doing phonetics by computer (version 5.0. 32). retrieved august 12, 2008.

Chen, M.Y., & W.S-Y. Wang. 1975. Sound change: actuation and implementation. *Language* 255–281.

Cheng, C.C. 1973. *A synchronic phonology of Mandarin Chinese*. Mouton The Hague.

Cheng, L.L.S., *et al.*. 1991. Feature geometry of vowels and co-occurrence restrictions in cantonese. In *WCCFL*, volume 9, 107–124.

Chitoran, I., L. Goldstein, & D. Byrd. 2002. Gestural overlap and recoverability: Articulatory evidence from georgian. *Laboratory phonology* 7.

Clayards, M. 2010. Using probability distributions to account for recognition of canonical and reduced word forms. In *LSA Annual Meeting Extended Abstracts*, number 0.

——, M.K. Tanenhaus, R.N. Aslin, & R.A. Jacobs. 2008. Perception of speech reflects optimal use of probabilistic speech cues. *Cognition* 108.804–809.

Delattre, P. 1958. Unreleased velar plosives after back-rounded vowels. *Acoustical Society of America Journal* 30.581.

Delattre, P.C., A.M. Liberman, & F.S. Cooper. 1955. Acoustic loci and transitional cues for consonants. *Journal of the Acoustical Society of America* .

Dorman, M.F., M. Studdert-Kennedy, & L.J. Raphael. 1977. Stop-consonant recognition: Release bursts and formant transitions as functionally equivalent, context-dependent cues. *Attention, Perception, & Psychophysics* 22.109–122.

Fant, G. 1970. *Acoustic theory of speech production with calculations based on X-ray studies of Russian articulations*. Number 2. Mouton De Gruyter.

Flemming, E. 2001. Scalar and categorical phenomena in a unified model of phonetics and phonology. *Phonology* 18.7–44.

—— 2004. Contrast and perceptual distinctiveness. *Hayes, Kirchner and Steriade (eds.) Phonetically Based Phonology* 232–276.

Flemming, E.S. 2002. *Auditory representations in phonology*. Routledge.

Halle, M., G.W. Hughes, & J. Radley. 1957. Acoustic properties of stop consonants. *Journal of the Acoustical Society of America* .

Householder, F.W. 1956. Unreleased ptk in american english. *Halle, Lunt, McLean, van Schooneveld, For Roman Jakobson* 235–244.

Jun, J. 2004. Place assimilation. *Phonetically based phonology* p. 58.

Kawasaki-Fukumori, H. 1992. An acoustical basis for universal phonotactic constraints. *Language and Speech* 35.73–86.

Kewley-Port, D., D.B. Pisoni, M. Studdert-Kennedy, *et al.*. 1983. Perception of static and dynamic acoustic cues to place of articulation in initial stop consonants. *Journal of the Acoustical Society of America* 73.1779–1793.

Kochetov, A., & C.K. So. 2007. Place assimilation and phonetic grounding: a cross-linguistic perceptual study. *Phonology* 24.397–432.

Liberman, A.M., P.C. Delattre, & F.S. Cooper. 1958. Some cues for the distinction between voiced and voiceless stops in initial position. *Language and speech* 1.153–167.

Luce, R.D. 1959. Individual choice behavior.

—— 1963. Detection and recognition. *Handbook of mathematical psychology* 1.103–189.

Malécot, A. 1958. The role of releases in the identification of released final stops: a series of tape-cutting experiments. *Language* 370–380.

Ohala, J. 2003. Phonetics phonology dissimilation. *Recherche* 67.02.

Ohala, J.J. 1993. Sound change as nature's speech perception experiment. *Speech Communication* 13.155–161.

Ohala, M., & J.J. Ohala. 1998. Correlation between consonantal vc transitions and degree of perceptual confusion of place contrast in hindi. In *Fifth International Conference on Spoken Language Processing*.

Ohde, R.N., & K.N. Stevens. 1983. Effect of burst amplitude on the perception of stop consonant place of articulation. *The Journal of the Acoustical Society of America* 74.706.

Öhman, S.E.G. 1966. Coarticulation in vcv utterances: Spectrographic measurements. *The Journal of the Acoustical Society of America* 39.151.

Olive, J.P., A. Greenwood, & J.S. Coleman. 1993. *Acoustics of American English speech: a dynamic approach*. Springer.

Repp, B.H. 1983. Coarticulation in sequences of two nonhomorganic stop consonants: perceptual and acoustic evidence. *Journal of the Acoustical Society of America* 74.420–427.

Winitz, H., M.E. Scheib, & J.A. Reeds. 1972. Identification of stops and vowels for the burst portion of/p, t, k/isolated from conversational speech. *The Journal of the Acoustical Society of America* 51.1309.

Wright, R. 2004. A review of perceptual cues and cue robustness. *Phonetically based phonology* 34–57.