Voice assimilation in French obstruents: A gradient or a categorical process?

Pierre Hallé, Martine Adda-Decker

Voice assimilation in French obstruents: categorical or gradient?

Pierre A. Hallé, LPP (CNRS-Paris 3)
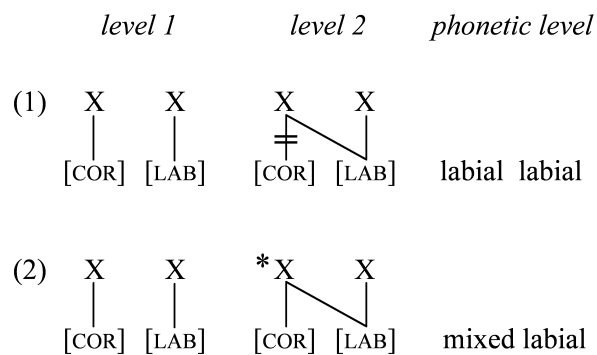Martine Adda-Decker, LIMSI (CNRS)

Abstract

This work contributes to the issue of categoricity versus gradiency in natural assimilations. We focused on voice-assimilation in French and started from the assumption that the main cue to obstruent voicing is glottal pulsing. We quantified glottal pulsing continuously with a single acoustic measure — the proportion in duration of voiced portion(s) within a consonant — which we call *v-ratio*. We used a large corpus of French radio and television speech to compute v-ratios for all the obstruents appearing in word-final to word-initial obstruent contacts. The results were analyzed in terms of v-ratio distributions, which were compared with theoretical distributions predicted by two contrasted hypotheses on the mechanisms of assimilation: categorical switch versus v-ratio shift. The comparisons strongly suggested that, although voicing itself can be incomplete, voice *assimilation* is essentially categorical in terms of v-ratio. We discuss this result in the light of recent perceptual data showing sensitivity to extremely subtle acoustic differences: secondary cues to voicing do not seem to follow the same pattern of categoricity as glottal pulsing.

1. Introduction

In one of his often cited papers, "The geometry of phonological features", Nick Clements wrote: "…there should be three common types of assimilation processes in the world's languages: TOTAL assimilation processes in which the spreading element A is a root node, PARTIAL assimilation processes in which A is a class node, and SINGLE-FEATURE assimilation processes in which A is a single feature." (Clements 1985: 231). An example of TOTAL assimilation would be *sweeb boy* for *sweet boy*: the last phoneme in *sweet* is substituted with the first in *boy*. But we know that place assimilation in English is only PARTIAL by Clements' definition: *sweet boy* becomes *sweep boy*, that is, only the place node is spreading, not the entire root node. Common to both types of assimilation,

however, spreading of a node from a source to a target segment (the assimilating and assimilated segments, respectively) is accompanied by the delinking of a "stray" node from the target segment. This part of the process is known as "stray erasure."

On that view of assimilation, assimilated segments retain nothing of the substituted node, whether assimilation is TOTAL or PARTIAL. At a phonetic level of description, this might entail that, for instance, /p/ in *sweep boy* is realized as a nonambiguous labial [p], with no traces of the underlying coronal place. This view of categorical assimilation is illustrated in (1). In (1) and (2), *level 1* corresponds to an abstract level of representation, before any contextual rule has applied, whereas *level 2* integrates the application of such rules. The *phonetic level* describes the output phonetically.

	level 1	*level 2*	*phonetic level*

(1)

X X X X

[COR] [LAB] [COR] [LAB]   labial labial

(2) *X X * X X

[COR] [LAB] [COR] [LAB]   mixed labial

(1) and (2) both assume that assimilation systematically happens when immediate context conditions are met. But on a slightly more complex elaboration of these descriptions, assimilation may or may not take place based on immediate context alone: assimilation conditioning is more complex. To make the issue simpler, however, we will assume that (1–2) are further qualified in that assimilation may not *always* occur.

Let us go back to the *sweet boy* example. By the categorical account in (1), assimilation, when it takes place, substitutes /p/ for /t/ at level 2, hence [p] to [t] at the phonetic level. Because both the intended sound /p/ and its context /b/ agree on place specification, no coarticulation for place is expected. When assimilation does not take place, /t/ is left unchanged and place coarticulation potentially affects phonetically the sounds in contact: [t] and [b]. Although such coarticulations may produce phonetically

intermediate sounds, the categorical nature of assimilation in (1) is not challenged.

By a gradient, incomplete view of assimilation, assimilated sounds are not so fully modified that they change category. They take on part of the assimilating context's phonetic characteristics yet retain part of their original characteristics, just like they are coarticulated. But incomplete assimilation, if this refers to an intentionally planned, active process, must entail something more than coarticulation albeit something less than complete, categorical assimilation. In our view, coarticulation cannot be such an active process (but see Flemming 1997) and coarticulation effects result from unintended biomechanical constraints.[1] In the sketch of incomplete assimilation shown in (2), the non-categorical nature of the process is therefore implemented at level 2 of the representation. Since (2) lacks delinking of the original specification, the process described is hardly acceptable phonologically (hence the '*' mark): It violates the basic premise that phonological representations be categorical. But how else can we formalize the notion of an assimilation process that is *intentionally* incomplete? Whatever the appropriate phonological formalism, the implication is clear at the phonetic level: When a sound is assimilated, its phonetic characteristics should always be intermediate between their original specification and that of the assimilating sound.

In section 2, "Gow versus Gaskell," we review the claim that assimilations occurring in natural speech are not categorical but gradient (in particular, Gow 2001, 2002; Gow and Im 2004). This claim is in line with the account of assimilation described in (2), whereas it is obviously contrary to that in (1). This latter account is *de facto* ruled out in the gradient view of assimilation. The claim that natural assimilations are incomplete has important consequences for our understanding of how the human system of speech processing copes with assimilation, that is, compensates for assimilation. Gow proposes a compensation mechanism, which he calls *feature cue parsing*, based on the premise that assimilations are *always* incomplete. We review briefly some recent counter-evidence to that view.

In section 3, "regressive voicing assimilation," we narrow in on voicing assimilation, which is the focus of the present chapter. We review phonetic evidence for incomplete assimilations and conclude that only distributional data could convincingly settle the ongoing debate.

In section 4, "corpus analyses," we present our own analyses of a large corpus of French broadcast speech, using the rough but robust index of voicedness that we call "voicing ratio" or *v-ratio*. The distributional data for v-ratio in assimilating contexts are compared with modeled distributions based on (a) the distributional data in non-assimilating contexts, and (b) the changes in distribution that assimilation imposes as predicted by categorical versus gradient accounts of assimilation mechanisms. Our modeling data suggest assimilation is more categorical than gradient in nature and that the phonetic basis for [±voice] is fully voiced versus not fully voiced. But these results must be tempered since we only looked at v-ratio. Although this index seems to capture most of the variation determining the perception of [+voice] vs. [−voice] in French, voicing is also cued by several other well documented characteristics.

Some of these characteristics are addressed in section 5, "subtle traces of voicelessness." In this section, we present a brief survey of a recent study (Snoeren, Segui, and Hallé 2008) suggesting that listeners can recover an original [−voice] specification even though v-ratio alone would indicate complete voicing. Traces of the original [−voice] seem to involve durational patterns as well as, to a lesser extent, microprosodic variation, and amplitude of glottal pulsing during consonant closure.

We conclude that the picture of voice assimilation in French is more complex than previously thought. Voice assimilation is largely categorical with respect to v-ratio. Yet, the "secondary cues" to voicing do not seem to undergo complete neutralization after voice assimilation. We discuss some possible ways to integrate such dissociation in a phonological description.

2. Gow versus Gaskell

A few years ago, David Gow published a series of studies on the issue of "compensation for assimilation." Previous studies had defended two different accounts of this perceptual device. On the "regressive inference" view defended by Gaskell and Marslen-Wilson, the right phonetic context licenses or not the recovery of the assimilated sound's underlying identity. For example, *leam* in *leam bacon* can be understood as an instance of (intended) "lean," whereas *leam* in *leam gammon* cannot (Gaskell and Marslen-Wilson 1996). In cases of ambiguity created by place assimilation (e.g., *a quick rum/run picks you up*), regressive inference does not help for recovering the intended word (here, *run* or *rum*) (Gaskell and Marslen-

Wilson 2001). On the "underspecification" view (Lahiri and Reetz 2010), [CORONAL] is, as a rule, left unspecified. Thus, /n/ in "lean," for instance, would not be specified for place and neither *leam* nor *lean* would mismatch the lexical representation of "lean."

Gow noted that the studies cited above had used deliberate, complete assimilations although natural assimilations, that is, produced in natural speech, are often incomplete. His "feature cue parsing" proposal relies on the assumption that natural assimilations are *always* incomplete. In potentially ambiguous utterances such as *right/ripe berries*, the assimilated form of *right* is neither [raɪp] nor [raɪt]: cues to [CORONAL] and to [LABIAL] coexist at the contact between words and are assigned to *right* and *berries*, respectively. Thus, by the feature cue parsing mechanism, ambiguity is avoided and at the same time, the intended word *right* is recovered, even if it is realized closer to [raɪp] than to [raɪt].

This elegant mechanism has been criticized, however, on several grounds. One concern is that compensation for assimilation would function identically across languages differing with respect to their specific phonological process(es) of assimilation. This point is controversial: Gow found positive evidence for compensation for assimilation with both Hungarian and American listeners tested on Hungarian voice-assimilated utterances such as *oros dinastia* 'Russian dynasty' (Gow and Im 2004); Darcy found language-specific compensations for assimilation with English and French listeners on language-specific place (English) and voice (French) assimilations (Darcy et al. 2009). Another concern is that feature cue parsing allows recovery of the assimilated consonant *only if* it retained traces, however subtle, of its underlying, original identity. Recent data from large corpora suggest that complete assimilations do occur (Dilley and Pitt 2007). More critically, Gaskell and Snoeren (2008) found that compensation for place assimilation also applies to word forms with no possible traces of the "recovered" place. For example, *rum* in *a quick rum picks you up*, intended as "rum" not "run," is interpreted as an instance of "run" more often than the same *rum* in *a quick rum does you good*. Gaskell and Snoeren interpreted their results in terms of learned statistical inference: listeners have experienced a significant number of ...*rum picks you up*... utterances in which unambiguously labial *rum* was a strongly assimilated form of "run," and therefore (mistakenly) interpret *rum* as an instance of "run" rather than "rum" more often in a labial than alveolar context. Whether or not this interpretation is correct, the feature cue parsing

mechanism proposed by Gow clearly fails to explain such occurrences of counterproductive compensation.

Compensation for assimilation can rely on something other than traces of the initially intended category. But are mechanisms exclusively relying on such traces ever needed? At this point, it is necessary to gauge the likelihood of encountering assimilations that leave traces of the intended category, that is, gradient assimilations. To this end, we focus on voicing assimilation.

3. Regressive voicing assimilation

There is little controversy about the phonetic substrate of voicing in French, contrary to, for example, English. French voicing is *mainly* cued by the presence/absence of glottal pulsing during the constriction of obstruent consonants, be they stops or fricatives. That voicing assimilation in French is regressive is also a matter of consensus. But its complete versus partial nature is debated. In the opinion of Grammont, assimilation is not complete when occurring between words, as in *une robe courte* 'a short dress': "…la cessation des vibrations glottales préparée pour le *c* commence dès le *b*, qui devient une occlusive sourde tout en restant une douce." ['…the interruption of glottal pulsing planned for the *c* already takes effect at the *b*, which becomes a voiceless stop yet remains a soft [consonant].'] (Grammont 1933: 186; also see Fouché 1969). In other words, voiceless and voiced stops are characterized by a "soft" versus "strong" quality (i.e., lenis versus fortis) in addition to the presence versus absence of glottal pulsing. The phonetic transcription [b̥] would capture rather well the idea that /b/ retains its "soft" quality before a voiceless obstruent. (And likewise, [p̬] would retain the "hard" quality of /p/ when voice-assimilated before a voiced obstruent). Grammont also notes that within-word assimilation is always complete, provided it does not follow schwa deletion as in *médecin* pronounced (in free variation) [medəsɛ̃], [metsɛ̃] or [mɛtsɛ̃]. For example, "obtenir" is pronounced *optenir* "avec un *p* fort et non avec un *b* sourd" ['with a strong *p* not a soft voiceless *b*']. Grammont explains this is because the phonetic context of *b* in *obtenir* cannot change, contrary to that of *d* in *médecin*. Put differently, the phonological form of the word "obtenir" is fixed at the surface level of representation: It is [ɔptənir].[2]

Grammont's account of voicing assimilation was left unquestioned until an aerodynamic and phonetic study by Rigault (1967). Rigault convincingly

showed that voicing assimilation was complete in both within-word and between-word assimilation situations. For example, he found that 'd' in either *médecin* [metsɛ̃] 'medical doctor', or *guide savant* [gitsavɑ̃] 'learned guide') was pronounced [t], with the same acoustic and aerodynamic characteristics as 't' in *fuite secrète* 'secret escape', and in stark contrast with 'd' pronounced [d] in a voiced context (e.g., *guide zoulou* 'Zulu guide'). Rigault concluded that the "soft" vs. "strong" distinction proposed by Grammont and elaborated by Fouché has no phonetic bases and moreover is questionable in terms of representational economy (Martinet 1955).

What can "graded" or partial assimilation mean at a phonetic level of description? For place assimilation, it clearly means that acoustic cues in the assimilated sound are intermediate between the initially intended sound and its assimilating context. For example, vowel-to-consonant formant transitions in assimilated *right* compared to plain (unassimilated) *ripe* and *right* would point to a stop intermediate between alveolar and labial (cf. Gow, 2002). For voice assimilation, we may expect that graded changes in voicing be signaled not only by intermediate cue values but also by incomplete change of the *temporal extension* of these cues. This is what Jansen (2004; also see Jansen and Toft 2002) reported to find in Hungarian regressive voice assimilation. He found that "voicing duration" in plosives and fricatives was the main acoustic parameter sensitive to assimilation but did not undergo complete changes that would neutralize the voicing contrast. Although the complete picture Jansen reported was complex, voicing duration — measured as the duration of glottal pulsing within the obstruents under scrutiny — varied substantially for most obstruents between a no-assimilation, baseline condition and an assimilation condition. For example, voicing duration for /k/ changed from 27 ms in a voiceless context (e.g., *va**k** találkozott*) to 53 ms in a voiced context (e.g., *va**k** darabolta*). Conversely, voicing duration for /g/ changed from 70 ms in a voiced context (e.g., *vé**g** dominál*) to 30 ms in a voiceless context (e.g., *vé**g** távolodik*). On average, assimilation did not lead to complete neutralization with respect to voicing duration. In particular, voice-assimilated /k/ did not change to a plain [g]. Interestingly, other cues to voicing, such as duration patterns (closure, release, preceding vowel) were less consistently sensitive to regressive assimilation. Jansen concluded that Hungarian regressive voice assimilation is governed by different processes at the same time: phonetic (coarticulation) and phonological (categorical switch) processes. Gow and Im (2004) reported similar phonetic data on

Hungarian voice assimilation in fricatives, using the same measurements as Jansen and Toft (2002). They found no assimilation for word-final voiced fricatives such as /z/ and incomplete assimilation for voiceless fricatives such as /s/ in *oros* ('Russian'). Note that, instead of voicing durations, Gow and Im reported latencies of glottal pulsing onset, which they called VOT. This implicitly entails they consistently found glottal pulsing in the *right* portion of the fricatives and absence of glottal pulsing in the *left* portion, contrary to the findings of Barry and Teifour (1999) on voice assimilation in Syrian Arabic. Our own data, reported in section 4.3, is in line with Barry and Teifour's. In Gow and Im's data, the mean VOT for voiceless fricatives in a voiced context is much longer than for plain voiceless fricatives but not as short as for plain voiced fricatives, suggesting incomplete assimilation. But behind the mean values suggesting incomplete assimilation, a qualitatively different reality may lie: complete assimilation in most occurrences, as observed by Rigault in French, and no assimilation at all in some other occurrences. That is, assimilation might actually be categorical, not gradient. To resolve the categorical vs. gradient issue, distributional data, not just means are therefore necessary. We expand on this point in the next section. For the moment, we simply note that categorical assimilation may appear as incomplete *on average* because it does not occur *all the time* but nevertheless is complete when it does occur.

One attempt at examining distributional data was made in Snoeren, Hallé, and Segui (2006) who measured "assimilation degrees" for voice-assimilated stops in French. They argued that voice-assimilation in French is gradient because about 45% of the assimilations they observed fell in the [20%, 80%] intermediate range of assimilation degree values. Such a pattern indeed suggests that assimilation might not be extreme in a substantial number of cases. The statement is, however, problematic for methodological reasons. For underlyingly voiceless stops, the "assimilation degree" in Snoeren et al.'s (2006) study is the percentage of voicing. This index has been used in other studies for either stops or fricatives (e.g., Burton and Robblee 1997) and is equivalent, in percentage, to the v-ratio we use (Hallé and Adda-Decker 2007). By this definition of assimilation degree, it is thus implicitly assumed that the percentage of voicing is zero for plain, unassimilated voiceless stops. But this latter assumption is rarely met: Voiceless obstruents seldom lack glottal pulsing completely (as we show in section 4.1). A no-assimilation baseline is therefore needed to estimate any "assimilation degree." Snoeren et al. (2006) did not measure such a baseline. For the opposite direction of assimilation — devoicing —

the lack of a baseline is less problematic. Indeed, voiced stops in a voiced context are almost always voiced throughout, that is, their percentage of voicing is very often 100%. The definition of the "assimilation degree" for underlyingly voiced stops — 100% minus percentage of voicing — is therefore rather plausible. But interpreting incomplete voicing remains problematic: How is the percentage of voicing distributed in plain, non-assimilated voiceless consonants? How does it compare with the distribution observed for assimilated consonants? Does the comparison suggest a continuous or a discrete assimilation process? To answer these questions, we undertook a distributional analysis of v-ratios on a large corpus of journalistic speech, comparing contexts licensing assimilation to no-assimilation contexts.

4. Corpus analyses

The corpus data under scrutiny have been described elsewhere (Hallé and Adda-Decker 2007). We first summarize the main aspects that have already been reported, then present new distributional analyses.

The initial corpus consisted of over 100 hours of speech from radio and television news with several levels of transcription aligned on the speech signal (phonemic, lexical, and morphosyntactic). About 100,000 speech passages with word-final to word-initial obstruent contacts (C1#C2) were extracted. A few examples with possible assimilation are given in (3).

(3) *avec des* /avɛk#de/
     neuf décembre /nœf#desɑ̃br/
     trouvent que /truv#kə/
     quinze chars /kɛ̃z#ʃar/[3]

Voicing ratios (v-ratios) were computed for both C1 and C2 in all the passages. V-ratio simply is the proportion of voicing, in duration, within an obstruent (see above). We focus here on the v-ratios on C1. As we discuss in section 5, the temporal extension of voicing within an obstruent is not the only cue to voicing but it is a primary one (cf. Jansen 2004). Moreover, v-ratio is relatively easy to measure in a large corpus, provided robust voiced-voiceless decision and alignment procedures are used.[4] Here, we relied on the automatic alignment system developed at LIMSI laboratories (Adda-Decker and Lamel 1999; Gauvain et al. 2005), and the F0 extraction cross-correlation algorithm implemented in Praat (Boersma 2001): Glottal

pulsing, that is, voicing was considered to occur wherever F0 could be computed.[5] We took the variation of v-ratio between the no-assimilation "control" situations and the potential assimilation situations as a measure of assimilation. Note that the control situations are all the more needed given that there is uncertainty with the measurements, as is unavoidable with automated procedures. The overall averaged results are shown in Table 1.

Table 1. v-ratio for C1 in control vs. assimilating contexts in all possible C1#C2 contacts (UV and V for voiceless and voiced, respectively).

type of contact	voicing assimilation			devoicing assimilation		
	UV-UV	UV-V	Δ	V-V	V-UV	Δ
stop–stop	.46	.74	.28	.91	.79	−.11
stop–fricative	.47	.74	.26	.89	.74	−.15
fricative–stop	.22	.68	.46	.85	.36	−.49
fricative–fricative	.35	.57	.22	.85	.54	−.31
means	.38	.68	.30	.88	.61	−.27

4.1. Baseline v-ratios

As can be seen in Table 1, the v-ratio of voiceless obstruents in a voiceless context is much larger than zero (between .22 and .47) and is larger for stops than fricatives (.46 vs. .28). In the case of stops preceded by a vowel, the closure portion overlaps with the voicing lag of the vowel. It usually begins with a short voiced portion of decreasing amplitude, which corresponds to the vowel offset. This makes the use of baseline voiceless-voiceless contacts necessary for quantifying assimilation. In the case of fricatives, the v-ratio observed might also reflect the voicing lag of a preceding vowel. But why is it smaller than for stops? In order to fully understand this pattern, we need to know where and how the glottal pulses occur within the closure portion of stops or in the entire constriction of fricatives. We address this point in the discussion of "voicing patterns."

The largest v-ratios are found for voiced obstruents in a voiced context: in average, .9 for stops and .85 for fricatives. The baseline v-ratio for V-V is thus much closer to the theoretical extreme (1) than the UV-UV baseline v-ratio is to the opposite extreme (0), in line with our previous comments.

4.2. v-ratios of assimilated obstruents

The change in v-ratio with respect to baseline when voiceless obstruents are followed by voiced obstruents is the largest for fricative–stop contacts. In average, it is about .27 for stops and .34 for fricatives. For voiced obstruents followed by voiceless obstruents, the change in v-ratio is rather modest for stops (–.13 in average) and about three times larger for fricatives (–.40 in average), the largest, again, in the case of fricative–stop contacts (–.49). Overall, then, voicing and devoicing assimilations induce more or less symmetrical changes in v-ratio for fricatives but not stops. For stops there is much less devoicing than voicing, confirming the earlier findings in Snoeren et al.'s (2006) study, although that study did not use appropriate baselines to estimate assimilation degrees.

The issue at stake in the present paper is whether assimilations are complete or not. The data in Table 1 seem to provide a possible answer to this question in that the v-ratios for assimilated C1s never reach those for plain C1s. For example, whereas v-ratio for plain voiceless C1s is .38, v-ratio for devoiced C1s is .61 (in average). Likewise, the v-ratios for plain voiced C1s and voice-assimilated, phonologically voiceless C1s are .88 and .68, respectively. Based on these averages, assimilations could be viewed as incomplete. Yet, as we already noted, intermediate averages may hide an all-or-none underlying reality: Assimilations may not occur all the time but nevertheless be complete when they do occur. It is of course also possible that assimilations, when they occur, are incomplete. Only distributional data can help decide which account is the most plausible. We present such data for v-ratio in sections 4.4 and 4.5 and argue that they support the former rather than the latter account.

Before we look at v-ratio distributions, however, we need to resolve the concern mentioned earlier as to what v-ratios tell us about voicing. Do changes in v-ratio reflect changes in the temporal extension of voicing — changes in the cumulated durations of glottal pulsation — or changes in the strength of glottal pulsation? Because our v-ratio data was based on F0 (or, alternatively, harmonicity) computation, they must be sensitive to signal amplitude. More specifically, a 0.5 v-ratio, for example, may reflect that periodicity indeed occurs on half of the segment under scrutiny, or that it occurs on the entire segment but with such intensity fluctuations that periodicity is detected only half of the time. In order to resolve this issue, we need to know where voicing occurs in segments.

4.3. v-patterns: Where does voicing lie?

When v-ratio indicates that a segment is not voiced throughout its entire duration, the question arises as to which portion or portions of the segment have been found to be voiced. In particular, we need to track situations in which voiced portions are scattered throughout the entire segment. Such situations would likely correspond to weak voicing but with full time extension rather than voicing with partial time extension. We considered four configurations in addition to fully voiced and fully voiceless: a single voiced portion located at the left edge, at the right edge, or in the middle of the segment, or scattered voicing with possible contact at one or both edges. We call these configurations "voicing patterns" or *v-patterns*.

Table 2. Percentages of v-patterns in C1 according to voicing contact for stops and fricatives (UV and V for voiceless and voiced, respectively).

C1 manner	contact	no voicing	left voicing	fully voiced	right voicing	scattered voicing
stop	UV-UV	15.6	63.6	14.4	2.2	4.2
	UV-V	9.5	17.5	63.0	3.7	6.3
	V-V	1.9	5.3	88.5	1.6	2.7
	V-UV	8.5	29.2	58.9	1.2	2.2
fricative	UV-UV	24.6	64.1	3.6	2.6	5.1
	UV-V	7.9	27.6	52.0	3.1	9.5
	V-V	1.5	16.0	77.1	0.7	4.6
	V-UV	10.7	74.7	10.8	0.6	3.2

Table 2 shows the frequency of occurrence of each v-pattern according to C1 manner and to voicing contact, with the two last configurations pooled together under "scattered voicing." The v-pattern data clearly show that scattered voicing is rather infrequent. Most of the time, voicing lies on the entire segment duration or on the left edge (about 83% of the time). This suggests that decreases/increases in v-ratio are not due to an overall weakening/strenghtening of voicing (e.g., in terms of glottal pulsing amplitude) but, rather, to a reduction/increase in the time extension of voicing. Obstruents generally have a single voiced portion starting from obstruent onset and extending to the right, possibly until obstruent offset. The non-zero baseline v-ratios found for voiceless obstruents must therefore measure preceding vowel voicing lag in most situations, for

fricatives as well as for stops. This resolves the concern expressed in 4.1. This finding is in stark contrast with the "VOT" data reported by Gow and Im (2004) for Hungarian fricatives, which seem to always begin with a voiceless portion and end with a voiced portion.[6]

Do these patterns tell us something about the discrete nature of assimilations? They suggest a partial exchange from "fully voiced" to "left" and "no voicing": For voiceless fricatives, these "left" and "no voicing" populations decrease uniformly, to the benefit of the "full voicing" population. We examine this point in more detail in the next section.

4.4. v-ratio distributions: Changes induced by voice context

When voiceless obstruents are followed by voiced obstruents, their v-ratio globally increases. But how does the v-ratio distribution change? Figure 1 shows the v-ratio distributions for the UV-UV (baseline) and UV-V (assimilation) conditions.
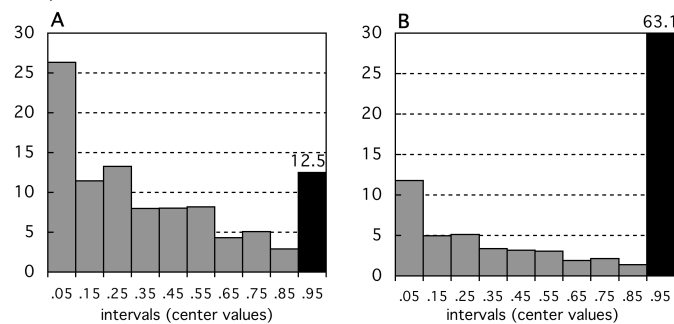

Figure 1. v-ratio distributions for C1 in the (A) UV-UV and (B) UV-V conditions; the black bar roughly corresponds to fully voiced C1s and is truncated in (B) to enhance detail in the [0, 0.9] range of v-ratios.

A striking aspect of these distributions is that, within the [0, 0.9] range, the v-ratio is not differently distributed for the baseline and assimilation conditions ($\chi^2(8) = 1.22$, $p = .996$). This suggests that voicing assimilations can be understood as a simple exchange between two categories, leaving unchanged the internal organization of each category. The two categories would be "fully voiced" (v-ratio > 0.9) on the one hand and "not fully voiced" (v-ratio in the [0, 0.9] range) on the other. However, the close similarity of the distributions within [0, 0.9] v-ratios between baseline and assimilation conditions does not hold across the board. It is found with

stops, for both voicing and devoicing assimilation and with fricatives for voicing assimilation but not clearly for devoicing assimilation. In the latter situation, v-ratios tend to decrease within the [0, 0.9] range. This suggests a different process than a simple exchange between the two categories proposed above or, perhaps, different v-ratio definitions of these categories. In section 4.5, we adopt a modeling approach to test further the categorical and graded accounts of voice assimilation.

4.5. v-ratio distributions: Modeling the changes caused by assimilation

By a categorical, discrete account of voice assimilation, there are two phonetically definable categories — voiced and voiceless — and the voice assimilation process simply is a *switch* from one category to the other. From a radical view of categorical assimilation, category switch always occurs in assimilation-licensing contexts. From a less radical view, there is either category switch or no category change. By the gradient, continuous account, assimilation can be viewed as a phonetic *shift* toward one of the two categories.[7] How can we model these two contrasting views of assimilation with respect to the single parameter examined so far — v-ratio? Figures 2 and 3 illustrate possible "shift" and "switch" scenarios, respectively, in the case of voicing assimilations. Devoicing assimilations are assumed to yield symmetrical scenarios. In these hypothetical scenarios, phonetic shift is modeled by an increase in v-ratio along the entire range of v-ratio values. This basically entails a rightward shift of the initial distribution. In particular, the leftmost peak that corresponds to the lowest v-ratios in the baseline condition (UV-UV) is shifted to the right by a constant amount. In our modeling, we have incorporated limit conditions and stochastic variation around a constant v-ratio shift d, as shown in (4).

The categorical switch scenario is modeled as a partial exchange between two posited categories — voiced and voiceless —, with no category-internal changes with respect to v-ratio. In such a model, the definition of the two categories is critical: The category boundary between [–voice] and [+voice] must be specified in the v-ratio dimension. The data discussed in 4.4 suggested a category boundary at a rather high v-ratio (cf. Figure 1). This is illustrated as 'boundary 2' in Figure 3. For the sake of comparison, a low v-ratio boundary is illustrated as 'boundary 1', hence two variants of the switch model: 'switch 1' and 'switch 2' (see Figure 3). The exchange model shown in (5) ensures that the within-category distributions of v-ratio are left unchanged after assimilation has applied.
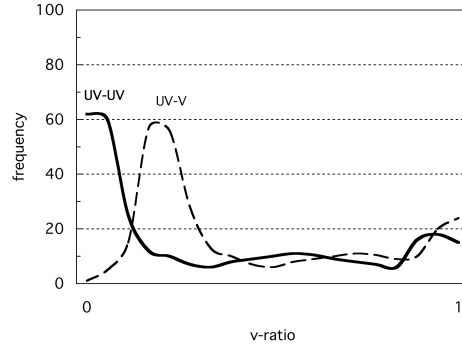
Figure 2. *shift model* for underlyingly voiceless obstruents: Made up of v-ratio distribution in voiceless context (plain line: baseline condition) and predicted distribution in voiced context (dashed line: assimilation).

(4) $f_a(v) = \alpha \times f_b(\max(1, \min(0, v + d + e(v))))$

(where *v* stands for v-ratio, $f_a(v)$ and $f_b(v)$ for the frequency of v-ratio *v* in the baseline (f_b) and assimilation (f_a) conditions; the scaling factor α ensures a constant cumulated frequency; *d* is the mean v-ratio shift.)
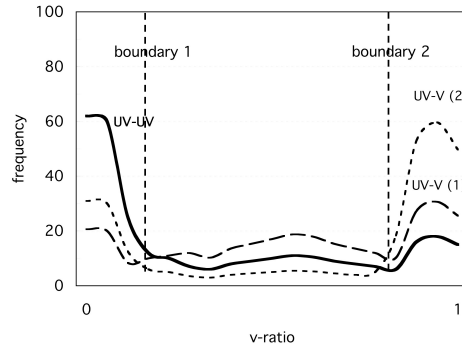

Figure 3. category *switch model* for underlyingly voiceless obstruents: Made up of v-ratio distribution for UV-UV (plain line: baseline) and predicted distributions for UV-V (assimilation) for two categorical boundaries (dashed line: boundary 1 at 0.2; dotted line: boundary 2 at 0.8).

(5) $f_a(v) = \begin{cases} f_b(v) \times r, & v < v_c \\ f_b(v) \times (1 - s), & v \geq v_c \end{cases}, \quad s = (r-1) \times \sum_{[0, v_c[} f_b(u) \Big/ \sum_{[v_c, 1]} f_b(u)$

(where v_c stands for a boundary v-ratio between the hypothetical voiceless

and voiced categories; *r* and *s* specify the amount of "exchange" between the two categories with no change in cumulated frequency.)

Which of these models best predicts the observed data? We computed, for each model and each underlying voicing, the v-ratio distribution in the assimilation condition predicted from the baseline distribution. This was done separately for stops and fricatives. The parameters *d* for the shift model (amount of shift)*,* and *r* and *s* for the switch models (amount of exchange) were estimated so that modeled and observed assimilations yield the same overall v-ratio. Figures 4A-C provides an illustration for the voicing assimilation of voiceless stops.
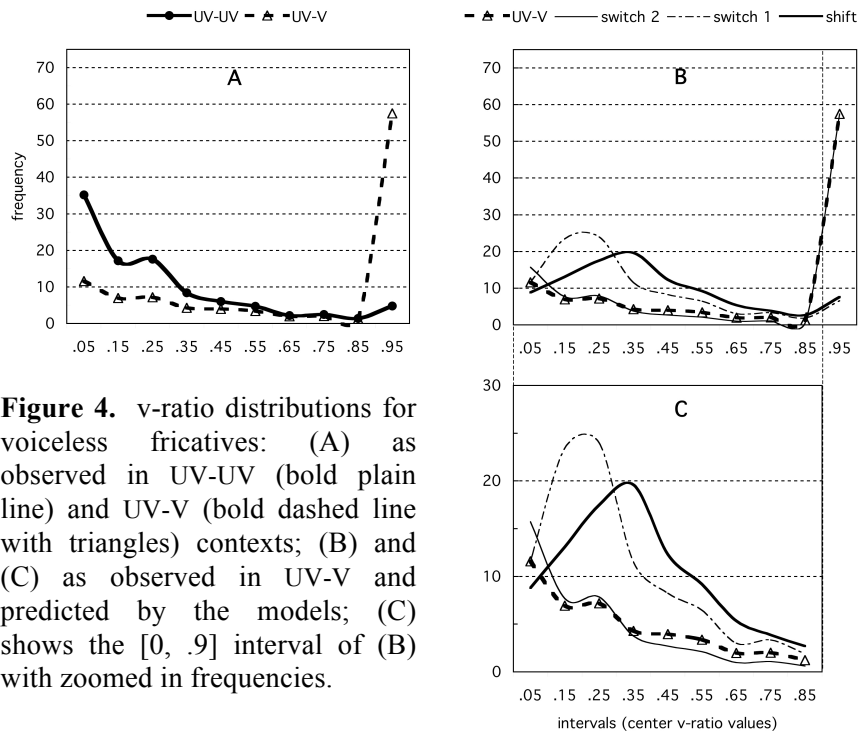

Figure 4. v-ratio distributions for voiceless fricatives: (A) as observed in UV-UV (bold plain line) and UV-V (bold dashed line with triangles) contexts; (B) and (C) as observed in UV-V and predicted by the models; (C) shows the [0, .9] interval of (B) with zoomed in frequencies.

To compare the models, root mean square deviations between modeled and observed distributions were computed. The results are shown in Table 3. The switch model with [0, .9] and].9, 1] v-ratios defining [−voice] and [+voice], respectively, clearly yields a better fit than the other two models, with a mean prediction error of about 2%. In detail, the adjustment is very good for all conditions excepted for devoicing in fricatives (5% error).

A closer inspection of the data reveals that assimilation does not affect within-category mean v-ratios except for this latter condition. Thus, for fricatives only, and for the devoicing direction of assimilation, there is a slight trend toward gradient assimilation in terms of v-ratio. Yet, for the most part, voice assimilations in French seem categorical in nature with respect to the voicing ratio parameter. Moreover, the data suggest a rather narrow phonetic definition of the categories. Voiced obstruents seem to be fully voiced in terms of v-ratio, whereas the voiceless category can be loosely specified as "not fully voiced." This points toward a default, unmarked [−voice] value of the [voice] feature. The marked value [+voice] is signaled phonetically by full voicing, with v-ratio = 1. But is v-ratio = 1 a sufficient condition for a segment to be [+voice]? Logically, that condition is necessary but perhaps not sufficient. The consequence in perception is that obstruents that are "not fully voiced" in terms of v-ratio (v-ratio < 1) should be perceived as [−voice], whereas obstruents that are fully voiced (v-ratio = 1) *may be* perceived as [+voice]. In the last section we examine recent perceptual data suggesting that v-ratio = 1 is not sufficient for the perceptual system to treat a segment as [+voice], at least in cases where the ambiguity between [+voice] and [−voice] cannot be resolved at the lexical level, that is, in cases where the surface form could correspond to different underlying forms, as in [sud] for either *soude* or *soute*.

Table 3. Adjustment scores (RMS prediction errors in %) for the three models examined; for the *switch* models, the [−voice] and [+voice] categories are defined by ranges of v-ratio variation: [0, .1] and].1, 1] for 'switch 1,' [0, .9] and].9, 1] for 'switch 2.'

model	V-UV		UV-V		average fit
	stop	fricative	stop	fricative	
shift	16.9	11.7	14.4	17.3	15.1
switch 1	6.6	18.6	16.0	18.0	14.8
switch 2	0.9	5.1	1.4	1.5	2.2

5. Subtle traces of voicelessness

In a recent paper, Snoeren, Segui, and Hallé (2008) used cross-modal associative priming to test for the effect of voice assimilation on lexical access. They used potentially ambiguous words such as *soute* /sut/ 'hold,' which is confusable with *soude* /sud/ 'soda' when strongly assimilated, that is, when pronounced close to [sud]. Other examples of minimal pairs for final consonant voicing included *trompe*, *jatte*, *bec*, *rite*, *bac*, *rate*, etc. (There are only about twenty such minimal pairs in French.) Snoeren et al. (2008) asked whether strongly voice-assimilated *soute* (pronounced close to [sud]) would activate the word "soute" not only at a phonological form level but further, at a lexical-conceptual level. In order to do so, they used natural assimilations of *soute* (pronounced in such utterances as *une soute bondée* 'a crammed compartment') that were very strongly assimilated (with v-ratio = 1). These word forms, extracted from the embedding utterances, were used as auditory primes in a cross-modal association priming experiment. For instance, "baggage" ('luggage') was paired with either *soute* pronounced [sud] or unrelated *gratte*. Other assimilated word forms such as *jupe* pronounced [ʒyb], which has no minimal pair for final consonant voicing, were used for comparison purposes: one possible outcome was indeed that only these unambiguous word forms would be accessed at a lexical-conceptual level. But the results clearly showed that unambiguous and potentially ambiguous word forms induced a comparable priming effect of about 40 ms. Hence, the lexical entry "soute" was activated by the strongly assimilated form [sud]. The critical question was of whether the word form [sud] for plain *soude* would *also* activate "soute." Indeed, spoken word recognition can be found to be relatively tolerant of mispronunciations (Bölte and Coenen 2000; Connine, Blasko, and Titone 1993; etc.). A second experiment showed that [sud] extracted from *une soude brute* 'a raw soda' did not prime "baggage" at all. The priming effect found with assimilated *soute* thus could not be due to form similarity with *soude*.

The only possible explanation of these data was that strongly assimilated forms (with v-ratio = 1), such as *soute* pronounced [sud], retain something of their underlying [−voice] specification. Snoeren et al. (2008) therefore set up to analyze the detailed acoustic characteristics of the assimilated stimuli they used. Table 4 summarizes the measurements that showed assimilated *soute* indeed retained something of /sut/. "V/(V+closure)"

summarizes the classic durational cues to voicing in obstruents: Longer preceding vowel and shorter closure for voiced obstruents, hence larger V/(V+closure). It seems virtually unaffected by voice assimilation. F0 on the preceding vowel offset seems to almost neutralize. Finally, the amplitude of glottal pulsing seems weaker for assimilated *soute* than for plain *soude*, suggesting that gradiency in voicing may be reflected not only by graded temporal extension, as found in several studies (Barry and Teifour 1999; Gow and Im 2004; Jansen and Toft 2002), but also by graded amplitude of glottal pulsing.

Table 4. Acoustic measurements of plain *soute* and *soude* (bold face) and of strongly assimilated *soute* (v-ratio=1) used in Snoeren et al. (2008).

	V-V *soude brute*	UV-V *soute bondée*	**UV-UV** *soute pleine*
v-ratio	1	1	0.38
V/(V+closure)	0.605	0.564	0.568
F0 at V offset	224 Hz	231 Hz	249 Hz
energy in closure	69.1 dB	67.4 dB	65.2 dB

To summarize, Snoeren et al.'s (2008) study clearly suggested that v-ratio cannot account entirely for the patterns of assimilation that are found in natural speech. Whereas some acoustic parameters seem to vary in an all-or-none manner — v-ratios change categorically, and durational parameters do not change — some others seem to vary in a graded manner (for example, amplitude of glottal pulsing). The picture of voicing assimilation is thus far more complex than previously thought.

6. Discussion

Let us summarize the observations that our corpus study made possible. The v-ratio means computed for the four voicing contacts suggested that voice-assimilated obstruents have intermediate v-ratios between those observed for their underlying voicing and those for the opposite voicing (Table 1). We showed that these means masked a different reality which could only be uncovered by examining distributions. Distributional data indeed suggested that assimilation takes place only part of the time but is complete, with respect to v-ratio, when it does take place. More precisely, two voicing categories may be defined phonetically (again, with respect to v-ratio): Full-voicing and partial-voicing. Assimilation, when it takes place,

is basically a switch between these two phonetic categories. How often does assimilation take place? A rough estimate can be obtained from the inspection, in Table 2, of the variation in frequency of the "fully voiced" pattern according to voicing contact. Since the frequency of this pattern increased by about 48% from UV-UV to UV-V contacts, for both stops and fricatives, we may infer that voicing assimilation takes place about 48% of the time. Likewise, devoicing assimilation seems to occur about 30% of the time for stops but 67% of the time for fricatives — an asymmetry already noted in 4.2. In section 5, we noted that secondary cues to voicing must remain unaffected or partially unaffected by assimilation since listeners can recover the intended voicing of fully voiced items, such as *soute* pronounced [sud]. Indeed, acoustic measurements revealed subtle differences between such items. In other words, apparently fully voice-assimilated forms retain traces of their underlying voicelessness. How can we reconcile the divergent observations for v-ratios and "secondary cues"?

Such dissociation between primary and secondary cues is reminiscent of the recent findings of Goldrick and Blumstein (2006) on tongue twisters inducing slips of the tongue. They found that when "k" was erroneously produced as [g] or "g" as [k], traces of the targeted consonant's VOT were found in the faulty productions. However, the "slip of the tongue" productions showed no traces of the targeted consonant in "local" secondary cues to voicing (F1 onset frequency, burst amplitude). As for the non-local cue examined — the following vowel's duration — it was faithful to the targeted consonant. For example, erroneous [k]s for targeted "g"s had a slightly shorter VOT than [k]s for plain /k/s, but had F1 and burst characteristics typical of /k/s, and maintained the long following vowel duration observed for plain /g/s.[8] (Symmetrical patterns obtained in the case of erroneous [g]s for targeted "k"s.) Goldrick and Blumstein claimed their data supported a "cascade" mechanism translating phonological planning into articulatory implementation: both the targeted and the slipped segment's representations were activated during phonological planning, resulting in a mix of both during articulation implementation. They also found evidence of cascading activations between the posited lexical level (or "lexeme selection") and phonological planning, all this supporting a cascade processing architecture across the board.

The assimilation data can be analyzed within the same framework of speech production planning and articulation (Levelt 2002). Following

lexical selection, phonological planning may proceed in several steps: A first step may activate a canonical representation at *level 1* posited in (1-2); when words are assembled together, contextual phonological processes may apply, activating *level 2* representations in a subsequent step. Thus, similar to the tongue twister situation, the voice assimilation process (or, more generally, any phonological alternation process) entails the activation of several phonological representations and representation levels, cascading to the articulation implementation stage. Hence, the possible mixed articulation implementation. This, together with coarticulation effects, must contribute to phonetically mixed outputs. Like Goldrick and Blumstein (2006), Snoeren et al. (2008) found a dissociation of cues in the observed voice assimilations but at the same time, the observed patterns were quite different. In assimilating contexts, v-ratios changed categorically but F0 on the preceding vowel, and waveform amplitude during stop closure underwent incomplete change, whereas preceding vowel duration did not change at all (just like Goldrick and Blumstein's (2006) following vowel durations). Goldrick and Blumstein (2006) interpreted the observed dissociation of cues as revealing the role of subsyllabic assembly mechanisms in articulatory implementation. They regarded the fate of secondary cues as explained by a lesser perceptual motivation. But this explanation obviously lacks consistency: Why should some perceptually unimportant cues be completely neutralized and some others entirely maintained? We propose instead that the dissociation of cues is due to different time-courses of phonological planning and articulation implementation. For instance, the resistance to assimilation for flanking vowel duration might be due to an early step of metric/prosodic planning completed before the assimilation process switches the [voice] specification of the assimilated segment. In the same way, we might interpret the weaker glottal pulsing during closure in voicing assimilations than in plain voicing as due to a later-occurring specification of voicedness in the assimilation case. We are of course aware that these interpretations are for the time being quite speculative and that more specific research is necessary to address, for instance, the issue of timing within the phonological planning stage and its possible consequences for articulation implementation.

Before closing, let us examine briefly the classical articulatory phonology account of assimilations in terms of gestural overlap (cf., for example, Browman and Goldstein 1992). Whatever the gestures involved for voicing in French — plausibly, glottal opening-closing for voiceless obstruents and glottal critical adduction for voiced ones (see Best and Hallé 2010 for an

overview) — the gestural overlap account predicts that assimilations occur in perception rather than in production, and are all the more likely to occur that speech rate is fast, or prosodic conditioning entails increased overlap. In other words, according to standard articulatory phonology, no discrete modification of phonological specification ever occurs in "phonological alternations"; Gestural specifications are not deleted nor switched: gestures may only overlap and hide each other *in perception*, especially at fast rates. One might want to test for this contention in corpus data: Assimilation degree should be stronger at faster rates. We attempted to do this by separating the data into four C1#C2 duration ranges (from less than 120 ms to more than 240 ms) and found near-categorical assimilation across the board with respect to v-ratio. This leaves little leeway for variations in assimilation strength according to, for example, speech rate.

To conclude, our data seems more readily amenable to a discrete rather than graded account of voice assimilation. In the scenario we propose, the classic description of assimilation, as found in (1), applies within the phonological planning stage in speech production: level 1 /sut/ produces level 2 [sud]. This takes us back to the typology of phonological alternations offered by Nick Clements: Voice assimilation belongs to the SINGLE-FEATURE type. The qualifications we propose to his typology are twofold. First, we propose, following Goldrick and Blumstein (2006), that a cascading architecture characterizes the translation from phonological code to articulation: Both /sut/ and /sud/ feed articulation implementation. In that view, the assimilations that are incomplete at the phonetic level, either quantitatively for a single cue (e.g., v-ratio, amplitude of glottal pulsing) or in terms of dissociation between cues, reflect cascading translation from phonological planning to articulation implementation, with different time courses of activation/deactivation for different levels of representation. In other words, whereas the classic description of the assimilation process in (1) offers a static picture, we propose to consider the dynamics of its component parts. As a second qualification, we introduced the notion of *occurrence* in the application of a phonological process. Immediate context determines whether assimilation is applicable or not. Yet, it seems that the actual occurrence of assimilation requires further determinants. What determines whether assimilation takes place or not? This question is indeed open to future investigation on the licensing factors that might operate beyond immediate context.

Acknowledgments

Notes

[1] Coarticulation is viewed here as a mechanical consequence of temporal overlap in articulation between consecutive sounds (Fowler and Saltzman 1993; Browman and Goldstein 1990, 1992). Coarticulation occurs with vowels (Öhman 1966; Magen 1993) or tones (Abramson 1979; Xu 1994), and indeed with consonants, in all the situations whereby sounds in contact differ in some *phonetic* dimension.

[2] Should we consider the pronunciation [ɔptəniʁ] instead of [ɔbtəniʁ] for *obtenir* as a case of within-word voice assimilation? This is a matter of debate. From a synchronic point of view, we may argue that the lexical form of "obtenir" is simply stored as /ɔptənir/ and there is no phonological context around or within that word possibly licensing an alternation with the /ɔbtənir/ form. However, at the abstract morphophonemic level, *obtenir* contains the prefix {ob-}, hence the phoneme /b/. The fact that *obtenir* has a /p/ at a less abstract level can be captured by a transformation rule governing the alternation between /p/ and /b/ in {ob-}, that is, by an assimilation rule taking place between levels of representation. The case of *médecin* is different because its pronunciation can alternate between [medəsɛ̃] and [metsɛ̃] or [mɛtsɛ̃]. Interestingly, 'é' in *médecin* is pronounced [e] more often than [ɛ], although it should be [ɛ] in the closed syllable /mɛt/ of /mɛt.sɛ/. This deviant pronunciation is symptomatic of a morphophonemic level of representation in which 'é' is indeed /e/, as reflected in the surface forms of *médical*, *médicament*, etc.

[3] Note that place assimilation may additionally occur in this example (Niebuhr, Lancia, and Meunier 2008).

[4] A discussion of the reliability and precision of the measurements presented here falls out the scope of this paper. There are indeed potential shortcomings in any automatic alignment system as well as in any automatic decision on acoustic voicing. (Manual labeling and measurement procedures are not error free either.) But the analyses proved to produce rather consistent and homogeneous patterns of results, which is about all what is needed for the present study.

[5] We compared this voicing decision procedure with a procedure based on the harmonics-to-noise ratio (HNR: a measure of acoustic periodicity) exceeding a fixed threshold. We set this threshold to 0 dB, which corresponds to equal energy in the harmonics and in the noise. The two methods yielded similar patterns of results.

[6] The opposite pattern we observe for French is also contrary to the naive intuition about regressive voice assimilation that the *right edge* of C1 should be affected by a following C2 with a different underlying voicing.

[7] Similar ideas have been offered by Massaro an Cohen (1983) in a different context. They proposed a new test for categorical perception in which listeners had to rate stimuli of a /b/-/d/ continuum on a 1-5 scale, as 1 if they heard /b/ until 5 if they heard /d/. Categorical perception predicts that subjects' ratings to a given stimulus be distributed along the 1-5 scale as two modes centered on the extreme ratings 1 and 5, whereas continuous perception predicts a single mode centered on a rating value depending on the stimulus, from 1 for /b/s to 5 for /d/s. That is, continuous perception predicts a distributional *shift* from one stimulus to another, whereas categorical perception predicts a *switch* between the two modes 1 and 5.

[8] In this study, the speech materials were strictly controlled and "distributions" restricted to limited dispersion around mean values. In other words, virtually *all* observed slips had slightly non-canonic VOT values.

References

Abramson, Arthur S.
1979   The coarticulation of tones: An acoustic study of Thai. In T. Thongkum, P. Kullavanijaya, V. Panupong, and T. Tingsabadh (eds.), *Studies in Tai and Mon-Khmer Phonetic and Phonology in Honour of Eugénie J.A. Henderson*, 1-9. Bangkok: Chulalongkorn University Press.

Adda-Decker, Martine, and Lori Lamel
1999   Pronunciation variants across system configuration, language and speaking style. *Speech Communication* 29 (2-4): 83-98.

Barry, Martin, and Ryad Teifour
1999   Temporal patterns in Arabic voicing assimilation. In *Proceedings of the 14th International Congress of Phonetic Sciences*, 2429-2432.

Best, Catherine T., and Pierre A. Hallé
2010   Perception of initial obstruent voicing is influenced by gestural organization. *J. Phonetics* 38: 109-126

Bölte, Jens, and Else Coenen
2000   Domato primes paprika: Mismatching pseudowords activate semantic and phonological representations. In *Proceedings of the SWAP Conference*, 59-62. Nijmegen.

Boersma, Paul
2001   Praat, a system for doing phonetics by computer. *Glot International* 5 (9/10): 341-345.

Browman, Catherine, and Louis Goldstein
1990   Gestural specification using dynamically-defined articulatory structures. *J. Phonetics* 18: 299-320.

Browman, Catherine, and Louis Goldstein
1992   Articulatory phonology: an overview. *Phonetica* 49: 155-180.

Burton, Martha W., and Karen E. Robblee
1997   A phonetic analysis of voicing assimilation in Russian. *J. Phonetics* 25: 97-114.

Clements, Georges N.
1985   The geometry of phonological features. *Phonology Yearbook* 2: 225-252.

Connine, Cynthia, Dawn Blasko, and Debra Titone
1993   Do the beginnings of spoken words have a special status in auditory word recognition? *J. Memory and Language* 32: 193-210.

Darcy, Isabelle, Franck Ramus, Anne Christophe, Katherine Kinzler, and Emmanuel Dupoux
2009   Phonological knowledge in compensation for native and non-native assimilation. In Franck Kügler, Caroline Féry, and Ruben van de Vijver (eds.), *Variation and gradience in phonetics and phonology*, 265-310. Berlin: Mouton De Gruyter.

Dilley, Laura C., and Mark A. Pitt
2007   A study of regressive place assimilation in spontaneous speech and its implications for spoken word recognition. *J. Acoust. Soc. Am.* 122 (4): 2340–2353.

Flemming, Edward
1997   Phonetic detail in phonology: Evidence from assimilation and coarticulation. In K. Suzuki and D. Elzinga (eds.), *Southern Workshop on Optimality Theory: Features in OT*. Coyote Papers.

Fouché, Pierre
1969   *Traité de prononciation française*. Paris: Klincksieck.

Fowler, Carol A., and Elliot Saltzman
1993   Coordination and coarticulation in speech production. *Language and Speech* 36 (2, 3): 171-195.

Gaskell, Gareth, and William Marslen-Wilson
1996   Phonological variation and inference in lexical access. *J. Exp. Psychology: Human Perception and Performance* 22: 144-158.

Gaskell, Gareth, and William Marslen-Wilson
2001   Lexical ambiguity resolution and spoken word recognition: Bridging the gap. *Journal of Memory and Language* 44: 325-349.

Gaskell, Gareth, and Natalie Snoeren
2008   The impact of strong assimilation on the perception of connected speech. *J. Exp. Psychology: Human Perception and Performance* 34 (6): 1632–1647.

Gauvain, Jean-Luc, Gilles Adda, Martine Adda-Decker, Alexandre Allauzen, Véronique Gendner, Lori Lamel, and Holger Schwenk
2005   Where are we in transcribing French broadcast news? *In Proceedings of Interspeech'2005-Eurospeech*, 1665-1668.

Goldrick, Matthew, and Sheila Blumstein
2006   Cascading activation from phonological planning to articulatory processes: Evidence from tongue twisters. *Language and Cognitive Processes* 21 (6): 649-683.

Gow, David W.
2001   Assimilation and anticipation in continuous spoken word recognition. *Journal of Memory and Language* 24: 133-159.

Gow, David W.
2002   Does English coronal place assimilation create lexical ambiguity? *J. Exp. Psychology: Human Perception and Performance* 28: 163-179.

Gow, David W., and Aaron M. Im
2004   A cross-linguistic examination of assimilation context effects. *J. Memory and Language* 51: 279-296.

Grammont, Maurice
1933   *Traité de Phonétique*. Paris: Delagrave.

Hallé, Pierre A., and Martine Adda-Decker
2007   Voicing assimilation in journalistic speech. In *Proceedings of the 16th International Congress of Phonetic Sciences*, 493-496.

Jansen, Wouter
2004   Laryngeal contrast and phonetic voicing: a laboratory phonology approach to English, Hungarian, and Dutch. Ph. D. diss., University of Groningen.

Jansen, Wouter, and Zoë Toft
2002   On sounds that like to be paired (after all): An acoustic investigation of Hungarian voicing assimilation. *SOAS Working Papers in Linguistics* 12: 19–52.

Lahiri, Aditi, and Henning Reetz
2010   Distinctive features: Phonological underspecification in representation and processing. *J. Phonetics* 38: 44-59.

Levelt, Willem
2002   Phonological encoding in speech production. In Carlos Gussenhoven and Natasha Warner (eds.), *Papers in Laboratory Phonology VII*, 87-99. Berlin: Mouton De Gruyter.

Magen, Harriet S.
1993   The extent of vowel-to-vowel coarticulation in English. *J. Phonetics* 25: 187-205.

Martinet, André
1955   *Économie des changements phonétiques : traité de phonologie diachronique*. Berne: Francke.

Massaro, Dominic W., and Michael M. Cohen
1983   Categorical or continuous speech perception: a new test. *Speech Communication* 2: 15-35.

Niebuhr, Oliver, Leonardo Lancia, and Christine Meunier
2008   On place assimilation in French sibilant sequences. In Proceedings of the 8th International Seminar on Speech Production, 221-224.

Öhman, Sven E. G.
1966   Coarticulation in VCV utterances: spectrographic measurements. *J. Acoust. Soc. Am.* 39 (1): 51-168.

Rigault, André
1967   L'assimilation consonantique de sonorité en français: étude acoustique et perceptuelle. In B. Hála, M. Romportel, & P. Janota (eds.), *Proceedings of the 6th International Congress of Phonetic Sciences*, 763-766. Prague: Academia.

Snoeren, Natalie, Pierre A. Hallé, and Juan Segui
2006   A voice for the voiceless: Production and perception of assimilated stops in French. *J. Phonetics* 34 (2): 241-268.

Snoeren, Natalie, Juan Segui, and Pierre A. Hallé
2008   On the role of regular phonological variation in lexical access: Evidence from voice assimilation in French. *Cognition* 108 (2): B512-B521.

Xu, Yi
1994   Production and perception of coarticulated tones. *J. Acoust. Soc. Am.* 95 (4): 2240-2253.