HAL
open science

The processing of illegal consonant clusters: A case of perceptual assimilation?

Pierre Hallé, Juan Seguí, Uli Frauenfelder, Christine Meunier

To cite this version:

Pierre Hallé, Juan Seguí, Uli Frauenfelder, Christine Meunier. The processing of illegal consonant clusters: A case of perceptual assimilation?. Journal of Experimental Psychology, 1998. hal-01771628

HAL Id: hal-01771628

https://hal.science/hal-01771628v1

Submitted on 19 Apr 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

The processing of illegal consonant clusters: A case of perceptual assimilation?

Pierre A. Hallé and Juan Segui

Laboratoire de Psychologie Expérimentale, CNRS and Paris V

Uli Frauenfelder and Christine Meunier

Laboratoire de Psycholinguistique, University of Genève

Running Head: Perceptual assimilation of illegal clusters

Mailing address: P.A. Hallé, LPE, 28 rue Serpente, 75006 Paris, France

e-mail address: `hal@idf.ext.jussieu.fr`

ABSTRACT

This paper presents evidence from a variety of experimental tasks for a perceptual shift, or assimilation. Consonant clusters that are phonotactically illegal, albeit pronounceable in French, are perceived as legal clusters, close in place and manner of articulation. Specifically, word-initial /dl/ and /tl/ are identified as /gl/ and /kl/, respectively in phonetic decision experiments. Two phonemic gating experiments show that subjects generally judge short gates Ñwhich do not yet contain information concerning the second consonant /l/ Ñ as being dental plosives. However, as information concerning the /l/ becomes available in larger gates, a perceptual shift develops, in which the initial consonants are increasingly judged as being velar plosives. We propose to call this special kind of perceptual shift "contextual perceptual assimilation". A final phoneme monitoring experiment suggests that this assimilation takes place on-line during speech processing, but not without some extra temporal processing cost. These results provide evidence for the automatic integration of low-level phonetic information into a more abstract code determined by the native phonological system.

INTRODUCTION

The view that speech perception is determined by the native language sound system is well motivated and widely shared. Ontogenetically, there is a shift from universal to language-specific perceptual capacities: Although young infants seem to be initially equipped with "universal" capacities for processing speech sounds, language-specific capacities emerge in the second half of the first year of life, by 9 months or before (Jusczyk, Cutler, and Redanz, 1993; Jusczyk, Friederici, Wessels, Svenkerud, and Jusczyk, 1993; Jusczyk, Luce, and Charles-Luce, 1994; Kuhl, Williams, Lacerda, Stevens, and Lindblom, 1992; Polka and Werker, 1994). By 10-12 months, children loose some of their discrimination capacities for nonnative consonants contrasts that are not functional in the language they are learning, at least when the consonants involved can be heard as (or assimilated to) possible speech sounds (Best, 1994; Best, McRoberts, Lafleur, and Silverisenstadt, 1996; Best, McRoberts, and Sithole, 1988; Werker and Tees, 1984). At the end of the first year, then, a language-specific speech processing system is well on the way toward stabilization. The language specificity of the resulting adult system has been characterized in terms of "phonological filter" metaphor (see Troubetzkoy, 1939). The native language sound system acts as a phonological filter molding some nonnative phones into native phonemic categories, provided there is sufficient similarity. In other words, nonnative phones may assimilate to native phonemes[1] (cf. Best's Perceptual Assimilation Model: Best, 1994; also see Flege's Speech Learning Model: Flege, 1986, 1991). Moreover, the attuning to the native language system also results in language-specific procedures in processing speech. For example, different native languages seem to foster different segmentation procedures, and, perhaps, use of different prelexical units of representation (Cutler, Mehler, Norris, and Segui, 1986, 1992). These findings can be explained by assuming that listeners are biased toward interpreting any spoken utterance as being consistent with the organization of their native language and with its functional linguistic representations.

Most studies dedicated to the general issue of native language bias have taken a cross-linguistic and developmental approach. At the same time they have focused largely on the

perception of single phoneme contrasts. Although several cases of cross-linguistic phoneme assimilation have been reported, little attention has been paid to *sequences of phonemes*, and in particular, to consonant clusters. The notion of a phonological filter, however, should apply to sequences of phonemes as well as to singleton phonemes. In the early thirties, Polivanov (1931) already claimed that the bias to perceive foreign sounds according to one's native phonological system is not confined to singleton phonemes but also extends to groups of phonemes. For example, Japanese listeners listening to the English word "drama" hear /do.ra.ma/ (Polivanov, 1931), consistent with the fact that Japanese does not allow syllable-initial clusters[2]. Native language phonotactic constraints thus can bias the perception of nonnative clusters so that they fit into the framework of the native language phonological system.

The effect of native phonotactic constraints in speech processing can also be studied *within* a given language. On the phonological filter view, illegal but pronounceable clusters should tend to be assimilated to clusters that are legal, whenever there is sufficient similarity, just as nonnative singleton phones tend to be assimilated to native phonemes. Hence, examining the perception of illegal clusters within a given language is yet another approach to explore how native language sound systems determine the perception of speech sounds.

So far, the perception of illegal clusters has not received much attention. Brown and Hildum (1956) presented natural speech monosyllables to native English speakers. The monosyllables had a complex onset and corresponded to (rare) English words or nonwords with either a permissible or non-permissible cluster at their onset. Subjects were required to write down what they heard. Their responses were strongly influenced by their expectation of English words and even more by their implicit knowledge of English phonotactic constraints. There was a strong bias to misperceive illegal clusters. This was observed for phonetically naive subjects but also for subjects trained in "linguistics", who were explicitly instructed to write phonemic transcriptions of non-English combinations of phonemes. Similar phonotactic knowledge and bias toward permissible clusters are also found in young children. Messer (1967) presented children learning English (aged 3;3 to 4;3 years) with pairs of monosyllabic nonwords, one with an illegal (or very infrequent) consonant cluster onset, the

other with a legal onset. (In a few items, the onset was /J/, which is illegal in English in that position.) The children were asked to say which item of each pair sounded "more like a word" by orally reproducing it. Legal onset items were chosen by children more often than illegal onset items. They mispronounced illegal items more often than legal ones, and the mispronunciations always turned illegal clusters into permissible clusters. As suggested by Messer, these results indicate "a perceptual disposition ... to hear sounds that are 'possible'..." This is in line with the notion of a phonological filter causing the perceptual assimilation of phones sequences as well as of single phones.

Another kind of evidence for perceptual assimilation of illegal clusters comes from Massaro and Cohen's (1983) study. Subjects were presented with synthetic speech stimuli beginning with an obstruent+liquid cluster. The liquid belonged to a /r/-/l/ continuum. The category boundary between /l/ and /r/ was shifted (relative to a neutral situation in which the initial obstruent was /p/) such that more /r/ phonemes than /l/ phonemes were identified after /t/ and, conversely, more /l/ phonemes than /r/ phonemes were identified after /s/. By this "phonological context effect" (in the authors' formulation), subjects implicitly showed a bias to perceive legal clusters, namely /tr/ or /sl/, rather than */tl/ or */sr/. Subjects may also have perceived the cluster-initial consonant differently as a function of the acoustic quality of the liquid. In Massaro and Cohen's (1983) Experiment 3, both the initial consonant and the liquid were ambiguous (a /b/-/d/ continuum was combined with a /r/-/l/ continuum). The results suggested a perceptual bias to report /bl/, /br/, and /dr/ rather than */dl/. For the /l/ endpoint, there were fewer /dl/ judgments for the /d/ endpoint than /bl/ judgments for the /b/ endpoint. In that sense, the perceptual shift appeared to affect the initial consonant too. The important point is that listeners showed a bias toward hearing legal clusters, that is, clusters consistent with the organization of their native language phonology.

In French as in English, /dl/ and /tl/ clusters are not permissible in syllable-initial position, although they are pronounceable. How are they perceived? A first possibility is that these clusters are perceived as non-French combinations of phonemes, with the constituent phonemes correctly perceived. That is, /dl/ would sound foreign, but nevertheless be perceived as the combination of /d/ and /l/. No perceptual assimilation would occur at the

structural level of syllable onset. This seems to be the case for some illegal Dutch clusters that were used in a recent experiment by Praamstra and Levelt (1994). In a lexical decision task, nonwords with such illegal clusters in initial position led to faster "No" responses than did phonotactically regular nonwords. This difference is best understood by assuming that the illegal clusters were readily perceived as impossible sequences. In contrast to the legal cluster nonwords, further speech input was not necessary to make a lexical decision. Presumably, then, these illegal clusters were not misperceived. A second possibility, suggested by the production data in the Messer (1967) study and by the Massaro and Cohen (1983) study, is that certain illegal clusters are perceived as legal ones. This is an instance of structural perceptual assimilation: Individually, the phonemes /d/ and /l/ would not be misperceived, but their combination gives rise to the perception of a different sequence from the intended /d/ + /l/, for example, /d/ + /r/.

One reason why we might expect the perceptual assimilation of /dl/ and /tl/ to legal clusters is that the phonotactic constraint barring them from French is not motivated by a distributional regularity, such as the constraint that excludes nasal+plosive clusters. Rather, in French, /dl/ and /tl/ represent an accidental gap within the set of obstruant+liquid clusters (henceforth, OBLI clusters, following Dell's, 1995, terminology) and may conceivably be misperceived as some legal OBLI cluster. In contrast, the onset cluster /nt/ for example, is *structurally* impermissible in that there exists no legal cluster sharing the same structure to which it could assimilate.

Now, will /dl/ be perceived as /dr/, as Massaro and Cohen's (1983) results might suggest, or as some other cluster? In other words, we do not yet know whether one phone of the cluster is perceptually more stable than the other and if so, which phone is most likely to be misperceived. In French, /l/ and /r/ have very different phonetic values: /l/ is a "clear" alveolar [l], while /r/ is an uvular [¢], often produced with some degree of frication. Thus, the /l/ in /dl/ or /tl/ is not likely to be perceived as (French) /r/. Instead, we expect a shift in the perception of the initial consonant.

The present study was designed primarily to determine how /dl/ and /tl/ clusters in item-initial position are perceived by French listeners: (1) literally, as /d/ + /l/ or as /t/ + /l/, or

(2) assimilated to a permissible sequence. If /tl/ and /dl/ induce onset-level perceptual assimilation, the question arises of which sequence of phones they will assimilate to. These issues are addressed in an open response test, in phoneme identification experiments, as well as in gating experiments. If the results indicate that, indeed, a perceptual shift occurs, a second issue naturally arises as to the level of processing at which the assimilation occurs. Does the perceptual shift operate in an automatic (irrepressible) way, or is it a post-perceptual process? The possible misperception of illegal clusters may result from a conscious reanalysis of the constituents that are perceived, or from an automatic recomputation. In both cases, some additional perceptual difficulty Ñrelative to the straightforward perception of permissible clustersÑ can be expected. These issues, partly addressed in the gating experiments, are examined further in a on-line phoneme detection experiment.

Experiment 1: open responses to /dl/- or /tl/-initial nonwords.

Initially, we wanted to discover what French listeners hear when they are presented with items beginning with the illegal cluster /dl/ or /tl/, pronounced by a French speaker. More specifically, if there is a systematic bias to hear another sequence than the one produced by the speaker, would it affect the initial dental plosive consonant or, rather, the following liquid? A simple and straightforward approach to answering this question is to let subjects transcribe freely what they hear when presented with the clusters under scrutiny.

Method

Stimuli: Four nonwords with an initial /tl/ ('tlabdo', 'tlabod', 'tlobad', and 'tlobda') , and four with its voiced counterpart /dl/ ('dlapto', 'dlapot', 'dlopat' and 'dlopta') were used as test stimuli. In addition, 24 phonotactically legal nonword filler items were also prepared; two with a /br/ and /gr/ cluster in initial position; others with a single initial consonants and six filler items with a /pl/ or a /sl/ cluster in medial position. Disyllables rather than monosyllables were chosen because monosyllabic nonwords are rare in French, or are close

in form to words. Using monosyllables could thus induce unwanted lexical effects. The material was recorded by a male native speaker of French on a DAT tape recorder, then digitized (10 kHz, 16 bit resolution) and stored in computer files[3]. The spectrograms of the test stimuli were inspected to ensure that the onsets were not produced with an epenthetic schwa after the item-initial dental plosive, turning the illegal "clusters" into legal sequences /dWl/ or /tWl/. No trace of a vocalic portion could be detected.

Subjects: Eighteen French subjects, native speakers of Parisian French, participated voluntarily in this experiment. They ranged between 21 and 30 years in age. All of them were phonetically naive and none reported any hearing deficit.

Procedure: The 32 nonwords were presented to the subjects in a randomized order with an ISI of 6 seconds. Subjects were tested individually in a sound attenuated booth, and received the speech material through Sennheiser headphones. After the presentation of each item, subjects wrote down what they heard, using a free style of transcription. These naive transcriptions were then inspected and recoded using a standard phonological transcription for French. When necessary, subjects were asked to clarify their responses.

Results and discussion

The outcome of this experiment was clear-cut: The /dl/ and /tl/ items were transcribed with a /gl/ or /kl/ cluster 85.4% of the time (*SD* = 7.2, range [72.2, 94.4]), with a /dl/ or /tl/ cluster only 13.2% of the time, with a /pl/ cluster 1.4% of the time (see Table 1)[4]. In contrast, the other consonantal word onsets, either singletons or clusters, were transcribed correctly 100% of the time. So were the word-medial clusters /sl/ and /pl/, except on a few occasions (6%) when subjects inserted an "e" in the cluster (e.g., "jocelaire" for [JOslE¢]). Clearly, subjects transcribed the auditorily presented /dl/ and /tl/ clusters more often as a combination of letters corresponding to /gl/ or /kl/ than as "dl" or "tl" (by-subject: $t(17) = 6.65$, by-item: $t(7) = 14.3$, both $p < .0001$), and never as "dr" or "tr". Is this outcome reflecting unwanted

lexical effects? None of the reported forms were close to a word. Moreover, in the context of -/a/ or -/o/ as well as other vowels, /gl/ and /kl/ are not more frequent word-initial clusters than /dr/, /tr/, /bl/, or /pl/[5]. The results thus suggest that, regardless of possible lexical influences, /dl/ and /tl/ assimilate respectively to /gl/ and /kl/, both legal clusters, apparently supporting the prediction that the perception of the /l/ segment in the cluster is more stable than that of the initial plosive consonant.

Table 1 about here

The high percentages of dental-to-velar "confusions" may be due to other reasons than perceptual assimilation per se. First, the /dl/ and /tl/ test items may have had an intrinsic velar quality, that is, velar rather than dental cues to place of articulation were perhaps present in the speech signal. This was possibly due to some difficulty that a native speaker of French may have in clearly articulating a speech item with a /dl/ or /tl/ initial cluster. It is also possible that a velar quality unavoidably arises from the coarticulation of a dental plosive and a French /l/. We attempted to check that possibility by means of an acoustic analysis. Although acoustic characteristics cannot provide unequivocal proof of dental rather than velar articulation, the spectral properties of the release burst have been found to be specific to the articulation place, at least to a certain extent (Halle, Hughes, and Radley, 1957; Stevens and Blumstein, 1978; Kewley-Port, 1983; also see Lahiri, Gewirth, and Blumstein, 1984): For dentals, energy is usually distributed at higher frequencies than for velars or labials. We therefore compared the spectra computed at the release burst for the /dl/ and /tl/ items and for matched /gl/ and /kl/ items with identical segmental context uttered by the same speaker (e.g., 'dlopat' *vs.* 'glopat'). We then used the spectral center of gravity as a simple and objective measure to characterize those spectra[6]. The spectral center of gravity should lie at higher frequencies for dental than for velars in the same context. This was found for the dental and velar items under scrutiny: 2.62 *vs.* 1.64 kHz, $t(7) = 7.64$, $p < .0002$. There was no overlap, except for the item 'dlopta' whose burst spectral center of gravity (1.79 kHz) was the lowest. (The highest percentage of /gl/ responses, 94.4%, occurred for 'dlopta' but also for 'tlabod'.)

So, the measurements show that /dl/ and /tl/ items were more dental (in the sense of more fronted) than comparable /gl/ and /kl/ items. Figure 1 illustrates that difference; it also shows spectra from a /pl/ item, which, as expected, exhibits a burst spectrum centered on rather low frequencies. But is this demonstration conclusive? We could as well say that /dl/ and /tl/ items were *less velar* than /gl/ and /kl/ items with respect to the release burst spectrum. Moreover, maybe more subtle cues than spectral cues signalled velar articulation. So the discussion is not closed. However, the possibility of gross and extensive mispronunciations of the dental + /l/ items as velar + /l/ can be rejected. We will return to the issue of dental *versus* velar place of articulation in Experiment 3.

Figure 1 about here

There is another alternative to the perceptual assimilation account. Even if the cluster constituents were correctly perceived, listeners had time to (consciously) reconsider what they had heard and make transcriptions that were compatible with their implicit knowledge of phonotactic constraints. This tendency might have been reinforced in this task of transcribing the input by a bias toward spelling correctness. Many subjects were perhaps reluctant to write down a combination of letters which is not permissible in word-initial position. The overwhelming dental-to-velar shift that was obtained may thus be partly due not only to a tendency to report permissible phonemic sequences but also to subjects' knowledge of permissible graphemes in word-initial position. Indeed, subjects' transcriptions often reflected the idiosyncrasies of the orthography of French words rather than regular phonemic transcriptions. (e.g., /vobral/ was often transcribed "veaubrale".) This kind of observation points to an orthographic bias, but is difficult to quantify. Although about 15% of the items presented were transcribed with the anomalous orthography "dl" or "tl", the velar responses (about 85%) were mostly due to subjects who consistently avoided "dl" and "tl": There was a majority of "conservative" subjects (12 out 18) whose responses always respected regular orthography, but only a few less conservative subjects who produced "dl" and "tl" responses. This pattern of results is consistent with the orthographic bias explanation.

The next experiment was designed to test this interpretation further. We used a forced-choice identification task to make the subjects focus their attention on the item-initial consonant. Since this task does not require listeners to write down letters that could possibly form illegal sequences, the possibility that orthographic knowledge affects the responses is reduced. If a massive dental-to-velar shift is still obtained, it should reflect either the intrinsic velar quality of the item-initial plosive in the test stimuli (which was not apparent in the acoustic analysis), or a perceptual shift at the level of syllable onset.

Experiment 2: forced-choice identification of cluster-initial consonants

In this experiment, listeners identified item-initial consonants in a forced-choice procedure. Two tests were conducted. In the first, subjects had to choose among three consonants, with three places of articulation (dental, velar, and labial), but with voicing held constant. In the second, only dental and velar places of articulation were proposed, but either voiced or not, so that voicing confusions could be tested. With a forced choice response, it is unlikely that orthography would interfere with the response. Moreover, because listeners' attention was focused on the initial consonant, the tendency to modify the initial phonetic perception to make the stimuli phonologically permissible was presumably reduced.

Method

Stimuli: The 8 /dl/ or /tl/ test items used in the previous experiment were matched with similar nonwords with a legal initial cluster /dr/ or /tr/, /gl/ or /kl/, and /gr/ or /kr/, such as 'drapot', 'trabod', 'glapot', etc. These served as control items for which few or no perceptual confusions were expected. In the first test (three-fold forced choice), subjects received two sets of 36 items in which dental, labial, and velar initial consonants were equiprobable. All the items were disyllabic nonwords. In one set, the initial consonants were voiced (/b/-/d/-/g/ forced choice); in the other set, they were voiceless (/p/-/t/-/k/ forced choice). In the second test (four-fold forced choice), subjects received a single set of 32 items containing the 8 /dl/

or /tl/ test items and the 8 corresponding /gl/ or /kl/ control items, plus 16 filler items whose initial consonant was (equiprobably) /d/, /t/, /g/, or /k/. The speaker who recorded the material of the first experiment again recorded the new items, using the same apparatus as before.

Subjects: Twenty four subjects, students at Université Paris V, participated for course credit in the first test; sixteen students participated in the second test. All were native speakers of Parisian French without any reported hearing deficit.

Procedure: The subjects were tested individually in a sound attenuated booth. Stimuli were presented through Sennheiser headphones at a comfortable level, with a three-second ISI. Subjects were instructed to circle one of the three or four letters proposed for each stimulus (first test: "P, T, K", or "B, D, G"; second test: "D, T, G, K"). In the first test, subjects were assigned one of two orders of presentation.

Results and discussion

Results may be described in terms of phonetic "confusions". Only place of articulation confusions could occur in the first test while voicing confusions could also occur in the second.

In the first test, no confusion occurred for the velar-initial control items. For this test, the description of the results is thus limited to the confusions that occurred for dental-initial test items (/dl/ or /tl/ initial cluster) and control items (/dr/ or /tr/ initial cluster). Table 2 summarizes the results obtained in this first test: Dental-to-labial confusions were marginal, and dental-to-velar confusions occurred mainly for the test items. The percentages of confusions were analyzed according to the *order* of presentation, the *initial* consonant (/t/ or /d/), and the *cluster* type (illegal: /tl/ and /dl/, or legal: /tr/ and /dr/). Both the dental-to-velar and the dental-to-labial confusions (treated as a repeated measure) were examined. *Order* had no significant effect, $F_1(1,22) = 2.9, 1$, $F_2(1,24) = 2.8$, both p «\symbol» .1. Dental-to-

labial confusions were rare as compared to dental-to-velar confusions, $F_1(1,22) = 54.4$, $F_2(1,24) = 72.0$, both $p < .0001$. Many more dental-to-velar confusions occurred for illegal than for legal clusters (55.7% against 4.2% overall), $F_1(1,22) = 75.1$, $F_2(1,24) = 72.4$, both $p < .0001$. The percentage of these confusions was higher for /tl/ (63.5%) than for /dl/ (47.9%), but the difference was marginally significant in the by-subject analysis only, $F_1(1,22) = 4.0$, $p = .055$.

Table 2 and Figure 2 about here

Figure 2 shows the results obtained in the second test. In this test, both place and voicing confusions could occur. The /gl/- and /kl/-initial items served as controls. The pattern of dental-to-velar confusions for /tl/ and /dl/ items was quite similar to that observed in the first test: 50% /d/«\symbol»/g/ confusions, 71.9% /t/«\symbol»/k/ confusions. Place *and* voicing confusions occurred for the /tl/ items (12.5 % /t/«\symbol»/g/ confusions), but not for /d/«\symbol»/k/. Place confusions (dental-to-velar for test items, or velar-to-dental for control items) were examined according to the *cluster* type (illegal or legal) and the *initial* consonant voicedness; place confusions with and without voicing confusion were counted separately (treated as a repeated measure). Almost no place confusion occurred for control as compared to test items, $F_1(1,15) = 65.5$, $F_2(1,12) = 170.7$, both $p < .0001$. There were more place confusions for /tl/ than for /dl/ items when voicing confusion was ignored, $F_1(1,15) = 15.3$, $p = .0014$, $F_2(1,6) = 11.7$, $p = .014$. When limiting the comparison to /t/«\symbol»/k/ *vs.* /d/«\symbol»/g/ confusions, the difference between /tl/ and /dl/ items became non-significant in the by-subject analysis, $F_1(1,15) = 2.9$, $p = .11$, but was still marginally significant in the by-item analysis, $F_2(1,6) = 4.2$, $p = .08$.

This experiment had a more constraining design than the preceding one in order to focus subjects' attention on item-initial consonants rather than on whole-words and to reduce the possible effect of orthographic knowledge. The results obtained showed a smaller effect than in the first experiment, but still a very substantial dental-to-velar shift. This result suggests that some kind of perceptual assimilation occurs for /dl/- and /tl/-initial items. This

identification experiment probably gives a more realistic picture of the phenomenon than the open response experiment, where extra-perceptual factors such as orthographic knowledge conceivably exaggerated the size of the effect.

Let us return now to the question of the acoustic quality of the cluster-initial plosives in the illegal clusters used. The acoustic measures already exclude gross and extensive mispronunciations of the intended dentals as velars. As pointed out earlier, subtle cues to velar articulation may have not been revealed by these measures. In the next section we will study how the cluster-initial plosive consonants are *perceived* on the basis of partial acoustic information by means of a gating experiment.

Experiment 3: gating task with legal *versus* illegal onset clusters

The acoustic measures have shown that the /dl/ and /tl/ stimuli are more fronted than the matched /gl/ and /kl/ stimuli. Altogether, this makes it less likely that velar cues were present in /dl/ and /tl/. We need, however, to obtain a more precise idea of how the cluster-initial plosives are perceived in the absence of the following liquid context. Blumstein and Stevens (1980) have shown that very brief stimuli, as short as 10-20 ms, sampled from the release burst of (synthetic) CV syllables can be reliably identified for consonant place of articulation. Hence, subjects might be able to correctly identify the initial portion of the /dl/ and /tl/ clusters for place of articulation when they have been spliced out of the following context.

The forward gating technique can be used to examine how brief portions of the stimuli, sampled from the release burst are perceived. Moreover, if a perceptual shift does occur and develops as longer fragments are presented, the gating technique can reveal the "time course", as it were, of such a shift. Since we are investigating the phonetic perception of the illegal clusters /dl/ and /tl/, a variant of the gating paradigm must be used in which listeners transcribe what they hear in lieu of guessing a word (Grosjean, 1980; Warren and Marslen-Wilson, 1987). Indeed, if listeners were instructed to guess a word, they could not possibly report /dl/- or /tl/-initial words, and the test would not tap into the perceptual shift at

a sublexical level of processing. We therefore introduce here a "transcriptional" variant of the gating paradigm that we will call *phonemic* gating.

How do listeners interpret the place cues as increasingly longer fragments of the /dl/- and /tl/-initial items are presented? Suppose that listeners report velar rather than dental sounds as soon as a stable perception of place has emerged. This would show that cues to velar place of articulation are actually present in the speech signal, and explain the velar responses that were obtained in Experiments 1 and 2. In contrast, if listeners systematically report dental rather than velar sounds, at least for short fragments, we can reject the "velar quality" explanation. Listeners, however, may not consistently produce dental responses. For this reason we need to compare the illegal clusters /dl/ and /tl/ with matched legal clusters with either dental or velar initial plosive consonant: /dr/ and /tr/ or /gl/ and /kl/. Dental-initial legal clusters yield a baseline for how dental place of articulation is normally perceived, while velar-initial clusters yield a baseline for how velar place is normally perceived.

Experiment 3 used a successive format of presentation: In a given sequence, the same speech item (a nonsense word) was presented incrementally. Experiment 3 was comprised of two tests using different subjects: One group of subjects listened to /dl/ and /tl/ items and another group to matched /gl/, /kl/, /dr/, and /tr/ items. This design was intended to avoid possible influences, assimilatory or contrastive, of one set of items on the other.

Method

Stimuli: The experimental items in the first test consisted of the 8 /dl/- or /tl/-onset nonwords used in Experiment 1 and of 8 filler items, also nonwords. The second test used the same 8 filler items and 8 test items with /gl/, /kl/, /dr/, or /tr/ onsets, matched with those of the first test. Half of the possible matched items were used for each onset, in order to limit the length of the experiment so as not to go beyond the attentional capacity of the subjects. These items were: 'glapot', 'glopta', 'klabdo', and 'klobda'; 'drapto', 'dropat', 'trabod', and 'trobda'. (So, both vocalic contexts -/a/ and -/o/ where used for each context.). The speaker used for the

previous experiments recorded the new items appearing in this experiment, using the same apparatus.

The initial fragment, or gate, of each nonsense word contained the initial portion of the signal up to 10 ms after the stop release burst, whether that stop was prevoiced or not. The other fragments were made increasingly larger, using a 20 ms increment. The final 4 ms of each fragment was attenuated by a raised cosine function so that there were no perceivable "clicks" at the gate offset. The tenth and last fragment thus included 190 ms of signal after the release burst. This roughly corresponded to the first syllable in all items. Table 3 shows the average intervals from the release burst to the acoustic onsets of the liquid, the following vowel, and the following stop closure and release, for the three types of items: dental + /l/, velar + /l/, and dental +/r/. These intervals were measured on the basis of spectrograms.

Subjects The subjects, all native speakers of Parisian French, were aged 22 to 26 years; 28 subjects participated in the first test, and 22 in the second test, for course credit. The subjects were phonetically naive and free of any hearing difficulties.

Procedure: For both tests, subjects received a total of 160 stimuli (16 items x 10 gates) in a successive presentation format. For each item, 10 fragments of increasing size were presented, and subjects had to transcribe exactly what they heard on an answer sheet (using a free style of transcription) and to give a confidence rating for their transcription on a 1-to-5 scale. Importantly, listeners were required to transcribe what they heard rather than to guess a word, as in the classical gating task. In case they could not report any speech sound at all, which often happened for the very first gate, they just wrote down a question mark. Subjects were tested individually. After each subject had completed the test, the experimenter checked the written responses and whenever necessary, asked subjects to clarify the phonetic value of any unclear transcriptions. In each test, the 16 items tested were presented in two different orders.

Results and discussion

Figure 3.c shows the distribution of the subjects' gating responses for /dl/ and /tl/ items as a function of the place of articulation. It can be seen that the percentage of dental responses was much higher than that of velar responses at each gate. The percentage of dental responses reached a maximum at the fourth gate (83%), then gradually decreased to 67% at the last gate, whereas the velar responses symmetrically increased, reaching 32% at the last gate. The same pattern was observed for both orders of presentation.

In contrast, the identification of /gl/ and /kl/, and of /dr/ and /tr/ (Figures 3.a and 3.b) was increasingly accurate, reaching about 96% correct responses at the last gate. For /gl/ and /kl/, the percentage of velar responses reached about 86% at the fourth gate and then continued to increase. Intended velars were massively judged as velars from the early gates on and there was little trace of early dental judgments: The percentage of dental responses reached a maximum of 11% at the second gate and then faded away. This pattern of results eliminates a possible interpretation according to which the pattern obtained for /dl/ and /tl/ clusters is the result of *perceptual compensation*. This possibility is indeed suggested by the finding that ambiguous items on an English /d/-/g/ continuum have been judged as dental in the non-fronting context /ar/-, but as velar in the fronting context /al/- (Mann, 1980, 1986; Fowler, Best, and McRoberts, 1990). Accordingly, an initially high percentage of dental judgments should be obtained with fronted variants of velars, as expected in the context of a following alveolar liquid: Velar plosives should be sufficiently close to dentals to be judged as such before the subsequent segmental context is available and can be used to *compensate* for the initially perceived frontedness. But in our data, few initial dental judgments and shifts to velar judgments were observed for the /gl/ and /kl/ clusters, while this pattern was massively dominant for the /dl/ and /tl/ clusters. Therefore, the many early dental judgments, later shifting to velar judgments when /l/ is detected, must reflect a perceptual shift rather than a perceptual compensation phenomenon. For /dr/ and /tr/ items, the convergence toward *correct* (i.e., dental) responses was slower than for /gl/ and /kl/ items, reaching the 85% level at the sixth gate. This was perhaps related to the longer duration of /r/ as compared to /l/ (see Table 3). Still, the important point is that there was no decrease in the percentage of dental

responses for dental + /r/ items, just as there was no decrease of velar responses for velar + /l/ items. This difference with the dental + /l/ items shows that a perceptual shift was at work for illegal clusters, not for legal clusters. Since the control situations provided by the matched /gl/, /kl/, /dr/, and /tr/ items yielded sufficiently clear results Ñ/gl/ and /kl/ are perceived as velar onsets, /dr/ and /tr/ as dental onsetsÑ, the analyses reported below are limited to the items of interest: the /dl/- and /tl/-onset items.

Table 3, Figures 3.a to 3.d about here

The mean gate number at which /l/ was identified was 3.6 (*SD* = 0.7); the mean gate number at which the vowel following /l/ was reported (correctly or not) was 4.6 (*SD* = 0.6); finally, although the last gate did not reach the upcoming /b/ or /p/ consonant release burst (see Table 3), there were sufficient cues in the speech signal for a majority of subjects (an average 17.6 subjects out of 28) to detect a labial plosive (mean gate number 7.3, *SD* = 1.0).

Subjects' responses were analyzed according to *place* of articulation (dental, velar, or labial), to the *order* of presentation, and to the *gate* number. *Order* had no significant effect (both by-subject, $F_1(1,26) < 1$, and by-item, $F_2(1,42) < 1$). *Place* was highly significant, $F_1(2,52) = 92.2$, $F_2(2,28) = 98.2$, both $p < .0001$. Overall, there were many more dental than velar responses (73.6% against 20.4%: $F_1(1,26) = 49.8$, $F_2(1,14) = 55.4$, both $p < .0001$), while labial responses were negligible (2.1%). *Gate* was significant, $F_1(9,234) = 14.2$, $F_2(9,126) = 27.3$, both $p < .0001$. Indeed, there was substantial variation in the proportion of dental or velar responses across gates (Figure 3.c). The largest variation, however, seems to occur in the first few gates, where the percentage of dental responses increased from an initial 60% to 83% at gate number four. Was the decline following the fourth gate significant? This point was examined by limiting the analyses to gates 4 to 10. The effect of *gate* on dental responses was significant in this portion: $F_1(6,156) = 12.5$, $F_2(6,84) = 13.3$, both $p < .0001$. Moreover, the correlation between gate number and percentage of dental responses from the fourth until the tenth gate was significant (by item: $r(54) = -0.38$, $p = .004$; by subject: $r(194) = -0.21$, $p = .004$). The decrease in dental responses between gates 4 and 10 was thus

meaningful. It was significantly larger for /tl/- than for /dl/-items (by-item analysis: 25% against 6.3%, $t(6) = 3.99$, $p = .007$).

The individual contribution of the different items was estimated by collapsing all ten gates: In a by-subject analysis, an *item* effect could thus be examined in addition to *place* and *order*. Overall, *item* was marginally significant, $F_1(7,182) = 2.0$, $p = .062$. As is seen in Table 4, three items had led to fewer dental responses (thus more velar responses) than the others. These were 'dlopat', 'dlopta', and 'tlobda'. (Recall that the burst of [d] in 'dlopta' was spectrally the least dental of all dental + /l/ items.) Item variability was indeed to be expected since we used natural speech. However, even the "least dental" test item collected an average 53% of dental judgments across the ten gates.

Table 4 about here

To sum up, the results show that, overall, the initial stops of /dl/- and /tl/-items were largely identified as dental rather than velar. However, subjects revised their dental judgments and reported more and more velars from the fourth gate on. Still, the percentage of dental responses did not drop below 67%. So, it is unlikely that cues to velar articulation were present in the speech signal for subjects to report, for example, /glapOt/ for 'dlapot' in the preceding experiments.

The shift toward velar identification began to occur by the fourth gate, where /l/ and the following vowel were first identified. From this point on, dental judgments declined and were revised into velar judgments. This points to a perceptual shift which is contextual rather than simply to a misperception of the illegal cluster's first constituent. Although the trend for a dental-to-velar shift was significant, it led to many fewer velar responses than in the preceding experiments. At the last gate, however, the entire CCV portion of the test item was presented (e.g., /tla/ for 'tlabod' or 'tlabdo'), that is, sufficient information for a contextual velar shift to occur. So why was the rate of velar responses so low as compared to the first two experiments? One possibility is that the assimilation process at work, whatever its nature, must take time to unfold completely, and that the responses given to the gated stimuli

reflected an intermediate state of the process, not a post-perceptual decision. Another possibility is that the particular format of forward gating used in Experiment 3, the successive format of presentation, induced perserveration in subjects' responses (Walley, Michela, and Wood, 1995; but see Cotton and Grosjean, 1984): Perserveration in reporting dental sounds might have masked the strength of the dental-to-velar shift effect. Walley et al. (1995) have shown that a duration-blocked format of presentation avoids perseveration effects, since, in this format, successive stimuli are fragments of different (and unpredictable) items. We thus replicated the gating task of Experiment 3, using the same stimuli but with a duration-blocked format of presentation to test for the perseveration explanation of the low rate of velar shifts that obtained in this experiment, even when the full liquid+vowel context was presented.

Experiment 4: gating task with stimuli blocked by duration

This experiment differed from Experiment 3 only in the format of presentation. Items were blocked by their duration so that response perserveration could not possibly bias the results. In particular, the percentage of dental responses at the last gate, which might have been inflated in Experiment 3, could become lower now. A "control" test using matched /gl/ and /kl/ items was no longer necessary since the dental nature of the /dl/ and /tl/ items had been sufficiently established in Experiment 3. The issue at stake was now whether or not a clear and abrupt shift from dental to velar judgments would be revealed using the duration-blocked format of presentation.

Method

Stimuli: The stimuli were the same as in the first test of Experiment 3.

Subjects: Twenty six subjects, aged 22 to 28 years, native speakers of Geneva French participated for course credit in the experiment. We did not expect significant differences

with Parisian French for the present experiment, since the two dialects mainly differ with respect to prosodic dimensions. None of the subjects reported hearing problems.

Procedure: The stimuli presented were blocked by duration. Each block contained 16 stimuli corresponding to the same gate number for each item; item order was randomly changed from one block to the next; successive blocks corresponded to increasingly larger gates.

Results and discussion

The results obtained were essentially the same as in the previous gating experiment. The percentages of dental, velar, and other transcriptions for the /dl/ and /tl/ items are shown in Figure 3.d. The percentage of dental responses reached a maximum at the fifth gate (91%), then gradually decreased (to 74.5% at the last gate) while velar responses symmetrically increased, reaching 24% at the last gate.

The mean gate number where /l/ was identified was 3.6 (SD = 0.8); the mean gate number where a vowel following /l/ was reported (correct or not) was 4.5 (SD = 0.8); finally, a few subjects (an average 6.6 out of 26) were able to detect a labial plosive in final position (mean gate number 8.5, SD = 0.9).

Analyses similar to those for Experiment 3 were performed, except that there was no *order* factor. In the analysis of subjects' responses according to *place* and to *gate*, the same effects as in Experiment 3 were found (*place*: $F_1(2,50) = 126.7$, $F_2(2,14) = 382.5$, both $p < .0001$; *gate*: $F_1(9,225) = 2.9$, $p = .0033$, $F_2(9,63) = 9.6$, $p < .0001$). Dental responses were much more frequent than velar responses (80.2% against 17.2%: $F_1(1,25) = 76.2$, $F_2(1,7) = 232.9$, both $p < .0001$). Labial responses were negligible (1.7%). The decline following the fifth gate was significant, as shown by analyses restricted to gates 5 to 10. The effect of *gate* on dental responses was significant in this portion: $F_1(5,125) = 7.4$, $F_2(5,35) = 14.7$, both $p < .0001$. The correlation between gate number and percentage of dental responses, from the fifth until the tenth gate, was significant, $r(46) = -0.68$, $p < .0001$ (by item), $r(154) = -0.21$, p

= .008 (by subject). The decrease in dental responses was significantly larger for /tl/- than for /dl/-items (23% against 9.6%, $t(6) = 2.94$, $p = .026$).

As shown in Table 4, between-item differences paralleled those found in Experiment 3. Again, 'dlopat', 'dlopta', and 'tlobda' led to somewhat fewer dental responses.

In sum, the results were similar to those obtained in the first gating task. The percentage of dental responses at the tenth gate was not smaller in Experiment 4 (74.5%) than in Experiment 3 (67%). Perserveration effects thus cannot explain the high percentage of dental responses in the gating tasks relative to those obtained in experiments 1 and 2.

Experiments 3 and 4 clearly demonstrated that the initial consonants in the /dl/- and /tl/-initial items did not have an intrinsic velar quality. Once a stable perception of place of articulation emerged (by the fourth gate), it was massively "dental" rather than "velar". In fact, even when most of the first syllable was presented, the rate of dental-to-velar shifts did not exceed 33%, whereas an overwhelming dental-to-velar shift obtained in the first two experiments. Aside from the difference in the size of the assimilation effect, the identification and gating results were strikingly similar in detail. In particular, the higher rate of confusions for /tl/ than for /dl/ in the identification tasks was paralleled by more frequent revisions from dental to velar judgment for /tl/ than for /dl/ in the gating tasks.

Was the modest size of the velar shift observed in the gating tasks due to the limited amount of information that was presented? At the last gate, the entire word-initial CCV portion, that is, a full acoustic syllable was presented. Since the velar shifts began to occur around the fourth/fifth gate, just after the liquid context was identified, it is not likely that more acoustic information was needed for the assimilation process to build up fully; rather, more time was needed. The responses to gated stimuli were thus reflecting the output of the assimilation process before its completion, at a point of time which, presumably, corresponded roughly to the endpoint of each fragment. The responses could have reflected a later state of the process at work, since the phonemic gating task imposed only a limited time pressure. One possible reason why they did not is that the gating task induced a kind of *a minima* strategy: The stimuli presented earliest in a session were so short, that is, so poor in phonetic information, that they encouraged responses based on low-level, straightforward,

phonetic perception rather than on phonological knowledge at a decisional stage[7]. Indeed, the subjects frequently continued reporting the impermissible sequences "dl" or "tl" well after the available context could influence them to reject such transcriptions. We will thus assume that the proportions of dental or velar responses in the gating tasks reflected the time course of the assimilation process. We will turn now to the nature of this process.

The gating data fit well with the notion of a "contextual" perceptual assimilation beginning to emerge as soon as listeners identify the liquid segment /l/, which, combined with the previously identified plosive, makes up an illegal cluster. But what kind of process caused this contextual assimilation?

One possible account is in terms of the unfolding of competing activations for at least two levels of prelexical representational units: single phones and syllable onsets. Dental plosives are initially (i.e., for the shortest fragments presented) more strongly activated than velars, and velars more activated than labials (for acoustic/articulatory reasons). However, as soon as the bottom-up information begins to activate the following /l/, the activation of velar plosives begins to increase at the expense of the dental plosives, presumably due to top down feed-back from onset representations favoring velars over dentals: Whereas /tl/ or /dl/ syllable onsets are in practical terms ruled out by a low phonotactic probability, /kl/ and /gl/ are not and can activate in turn the velar consonants. Labial plosives might conceivably benefit from a similar feed-back activation, but cannot overcome their initial handicap relative to velars. By this account, the decline of dental responses and the increase of velar responses that develops as longer fragments are presented reflect the time course of phone-level activations.

Another account of the data is provided by non-interactive models such as the FLMP model (Massaro and Oden, 1978; Massaro and Cohen, 1993). The FLMP model would state that the "dentalness" as opposed to the "velarness" or "labialness" of the item-initial plosives is judged from the evaluation and integration of two independent sources of information: the plosive itself, and its phonological context. The first source provides support for dentalness to the extent that the acoustic features of the plosive agree with a prototype /d/ or /t/; for the /dl/ and /tl/ test items, this source provides a lesser support for velarness and a very weak

support for labialness. The second source (the #_/l/ context) provides independent evidence for velarness or labialness and not for dentalness, supporting only the alternatives that are phonotactically compatible with the following /l/ context (as well as with the word-initial position). A relative-goodness rule applies at the decision stage, yielding quantitative predictions for the various alternatives. Also, both sources of information are processed over time. Conceivably, the contextual information occurs later in items such as /dlapto/. The FLMP could account for this by weighting the various amounts of support by a negatively accelerated grow function of processing time and introducing a fixed delay for the arrival of right context information (see Massaro and Oden, 1980, 1995). Appropriate parameters (6 degrees of 'static' support for three alternatives x two sources of information; rate(s) of information processing, and time delay(s)) could probably produce predictions of the dental, velar, and labial responses over time. Note that evaluating the support provided by phonological context (being in word-initial position and followed by an /l/) for each type of response is not basically different from evaluating the plausibilities of the corresponding clusters as syllable onsets, or, from assigning them initial levels of activation. The latter account is, perhaps, more appealing in that prelexical representational units are explicitly posited and avoids the underspecified notion of contextual source of information.

In sum, the experimental data do not allow choosing among the main competing theories. On both interactive and independent views, though, the perceptual shift observed here seems to engage automatic processes rather than post-perceptual decision. However, more evidence is needed to conclude that this perceptual shift results from an automatic process. An on-line task, like phoneme monitoring, could well provide such evidence.

Experiment 5: detection of cluster-initial consonants

Let us suppose that perceptual assimilation occurs in the on-line processing of speech as reflected by a task such as phoneme monitoring. Subjects should often miss dental targets in the /dl/ and /tl/ carrier items and, conversely, should incorrectly detect velar targets in these same items. Such a pattern of on-line misses and false positives would be analogous to the

dental-to-velar confusion data that were obtained in off-line tasks. Detection latencies of the velar targets in the dental + /l/ items could be longer than those for the same targets in the matched velar + /l/ items. Such a result would suggest an additional processing cost attributable either to a reanalysis of the phonetic input, or, in terms of competing activations, to a delay required for the velar activation to overcome the initial dental activation.

In this experiment, the same /dl/- and /tl/-items were used as in the gating and identification tasks, but this time in a generalized phoneme monitoring task (cf. Frauenfelder and Segui, 1989, for discussion of this technique). In addition to these items, subjects were also presented with the corresponding velar + /l/ as well as dental + /r/ legal sequences.

Method

Stimuli and design: The same 32 nonwords items that were used in the preceding experiments served as target-bearing carriers in this experiment, that is, eight legal (/tr/, /dr/) and eight illegal (/tl/, /dl/) dental+liquid clusters and eight legal (/kl/, /gl) and eight legal (/kr/, /gr/) velar+liquid clusters. These items were associated with four different lists, one for each phoneme target (/t/, /d/, /k/, /g/). Each list for velar targets contained four bisyllabic "foil" items with the corresponding illegal dental + /l/ clusters to test for possible incorrect detections resulting from perceptual assimilation. Similarly, the dental lists contained corresponding velar + /l/ foils to assess the probability of misperceptions in these cases.

Each list also contained items with targets in other positions and without targets. Four sets of 16 bisyllabic nonwords containing the respective targets in non-initial positions (medial and final) and two sets of 48 similar nonwords not containing the targets were constructed with a phonological structure similar to that of the target-bearing test items. Thus, each list contained the following items: eight item-initial target test items, 16 non-initial (medial and final) target-bearing fillers, four foils with the other place of articulation (velar or dental), and 48 non target-bearing items. Finally, one distractor list containing the same distribution of test and filler items was also constructed for a labial plosive target.

Four groups of subjects were each assigned to two test lists according to the phoneme targets (/t/ and /k/ targets *versus* /d/ and /g/ targets) and to the list presentation order (dental target first *vs.* velar target first). All subjects first received a practice list with 36 stimuli and a phoneme target not used in the test lists but appearing equiprobably in word initial, medial, or final position. The same distractor list with labial targets was inserted between the two test lists for all four groups. This distractor list served primarily to determine a common miss rate criterion across the phoneme monitoring lists. (Targets were unambiguous in this list, and a low miss rate was expected.) It served secondarily to break the monotony of the experimental sessions. The new stimuli were produced by the speaker who had recorded the material of the previous experiments.

Subjects: Fifty-two subjects, students at the Université Paris V, participated in the experiment for course credit. Four subjects did not reach the acceptance criterion (less than 50% misses) in the distractor list. None of the subjects reported any hearing difficulties.

Procedure and apparatus: Subjects were tested individually in a sound attenuated booth. They received the stimuli via Sennheiser headphones, at a comfortable listening level. They were told that they would hear nonwords and would have to make a speeded detection response to target phonemes by pushing a Morse key with their preferred hand. They were instructed that targets could occur anywhere in the carrier nonwords. Targets were specified in a booklet of cards upon which they were written as capital letters. The phonetic value of each letter had been explained and illustrated before the experiment. In particular, it was stressed that the letter 'G' stood for /g/ as in "gare" /gar/ not for /J/ as in "gilet" /JilE/. Subjects were orally warned (through the headphones) when the list changes occurred and were instructed at this point to switch to the next card for the subsequent target. The presentation of stimuli and of oral instructions, was controlled by a 486/66 microcomputer interfaced with an AD/DA OROS board (using 10 kHz sampling frequency and 16 bit resolution). The computer also collected reaction times, measured from the release burst of the targets. (All the targets were plosive consonants.) RTs were computed on the basis of the

number of samples that had been run through the DA converter at the moment the key was pressed.

Results and discussion

The percentages of misses and false positives are summarized in Figure 4. As can be seen, there were many false positive responses to velar targets and misses of dental targets for the illegal cluster dental + /l/ items. In contrast, false positives to dental targets as well as misses of velar targets were rare for the legal velar + /l/ items. This pattern of results provides strong support for the view that /dl/ and /tl/ are heard as or *assimilated to* /gl/ and /kl/ in on-line speech processing.

Figure 4 about here

The detection rates of dental and velar targets were analyzed according to *cluster* type (illegal /dl/ or /tl/ *vs.* legal /gl/ or /kl/) and experimental group (target *voicing* and *order* of presentation). Overall, neither *order* nor *voicing* reached significance (*order*: $F_1(1,44) < 1$, $F_2(1,24) < 1$; *voicing* : $F_1(1,44) = 1.9$, $p = .17$, $F_2(1,24) = 1.4$, $p = .24$). However, one group (voiced targets and dental target first) produced a comparatively high hit rate for the dental targets (68.8%), and low false positive rate for the velar target (66.7%) in the illegal dental + /l/ items: This group differed significantly from the other "voiced target" group in the illegal cluster condition, $F_1(1,44) = 5.9$, p = .018; $F_2(1,24) = 5.1$, p = .032. Despite the variability introduced by this group, the overall pattern of results was significant. In the illegal /dl/ or /tl/ items, false positives to velar targets were more frequent than hits to dental targets (79.7% *vs.* 33.7%), $F_1(1,44) = 78.3$, $F_2(1,12) = 51.6$, both $p < .0001$. Moreover, these false positives were about as frequent as the hits to velar targets in legal /gl/ or /kl/ items (79.7% *vs.* 81.8%), $F_1(1,44)$ or $F_2(1,24) < 1$. False positives to dental targets were rare in /gl/ and /kl/ items (5.2%). There was a non significant trend for the false positive rate to be higher for /k/ than /g/ in dental + /l/ items (83.3% *vs.* 76.1%), $F_1(1,44) = 1.1$, $p = .30$; $F_2(1,12) = 1.21$, $p = .29$.

To judge the effect of the cluster type on the detection of dentals, we compared directly the miss rates for dentals in illegal (/dl/ or /tl/) and legal (/dr/ or /tr/) clusters. The miss rate was higher in illegal than in legal clusters (68.8% *vs.* 25%), $F_1(1,44) = 155.9$, $F_2(1,24) = 43.2$, both $p < .0001$. There were more misses for /t/ than for /d/ target (81.2% *vs.* 68%) in illegal clusters, $F_1(1,44) = 11.9$, p = .0013; $F_2(1,12) = 7.9$, p = .015, but this was partly due to the group of subjects assigned to the detection of /d/ and list order dental-velar, where the miss rate of /d/ in /dl/ was unusually low (31.2%). Overall, the detection rate of dentals in the illegal clusters was so low that a RT analysis is meaningless. However, the prediction that dental targets are difficult to detect in illegal dental + /l/ clusters already received strong support from the miss rate data.

The second issue of interest is whether velar targets were more difficult to detect in the dental + /l/ than in the velar + /l/ clusters. Such difficulty was not apparent in the detection rates (79.7% false positives in dental clusters *vs*. 81.8% hits in velar clusters). It was, however, reflected in the RT data which could be compared directly given the relatively high and roughly comparable detection rates. Subjects were slower to detect, for example, /g/ in /dl/ (i.e., to false positive because of perceptual assimilation) than in /gl/ by about 34 ms (647 *vs.* 613 ms) as is seen in Figure 5. This difference was marginally significant in the by-item analysis, $F_2(1,24) = 3.24$, $p = .081$, but was reliable in the by-subject analysis, $F_1(1,46) = 6.4$, $p = .014$. The *target* factor (that is, /g/ *vs.* /k/) was not significant, both F_1 and $F_2 < 1$.

Figure 5 about here

Between-item differences tended to parallel those observed in the gating and off-line identification tasks: The perceptual assimilation effect was somewhat stronger with voiceless than with voiced dental + /l/ clusters. In the speeded detection task, this was apparent in the trend for the false positive rate to velars to be higher in /tl/ than in /dl/ items, and also, in the higher miss rate for dental detection in /tl/ than in /dl/.

A striking outcome of Experiment 5 was that the perceptual assimilation effect was larger than in the off-line identification task (80% false positive rate in detecting velars in

dental + /l/ clusters, against an average 61.5% dental-to-velar identification confusions). Because the assimilation effect was so massive in an on-line task, we can conclude that it is automatic and irrepressible, taking place at an unconscious level of processing. But the obvious question concerns the exact level of processing at which this perceptual assimilation takes place. We used homogeneous lists of nonwords so as to discourage a lexical strategy (see Frauenfelder, Segui, and Dijkstra, 1990, for a discussion of the list structure effects obtained, for example, in Marslen-Wilson, 1984), and consequently subjects presumably did not rely on lexical information. However, it also appears that subjects did not respond solely on the basis of phonetic-acoustic information. If this had been the case, they would presumably have identified the dentals in the dental + /l/ clusters like they did in the gated stimuli where dental perception in the early gates evolved toward velar perception. By analogy, subjects would have been faster to detect dentals than velars in illegal clusters, on the rare occasions that they did so. But they were not faster (642 ms *vs.* 647 ms for respectively 31% *vs.* 80% detection rates). Conceivably, then, the on-line detection of velar targets was made after the perceptual integration of the segments making up the cluster was completed. This suggests that perceptual assimilation occurred at a level of integration where a more abstract code than a phonetic code had been built, and where chunks larger than individual phonemes had been assembled. At the same time, this is an indication that phoneme detection is based on the perception of prelexical units larger than the phoneme, such as onsets. In contrast, as was proposed in the discussion of Experiment 4, the data for early gates seem to reflect phonetic perception before the perceptual integration of clusters has taken place.

GENERAL DISCUSSION

Experiments using a variety of different tasks have provided consistent evidence for the perceptual assimilation of illegal clusters (/tl/ and /dl/) into legal ones (/kl/ and /gl/). This is a robust phenomenon which is not due to misperception of the initial cluster constituent (the dental stop), but is induced by the combination of the two constituents making up the

syllable onset. We call it "contextual perceptual assimilation". It is in line with the notion of a language-specific phonological filter by which nonnative phones tend to be assimilated to native categories. The present findings show that this notion can be extended to explain perceptual assimilation *within* the native language of the listeners for sequences of phones such as consonant clusters. Certain phonotactically illegal but possible clusters assimilate to legal clusters. There is no reason to believe, however, that this phenomenon can be generalized to all illegal clusters. What we found for /dl/ and /tl/ in French does not necessarily apply to such impermissible clusters as /nt/ or /ms/. As was pointed to above, /dl/ and /tl/ sequences are predicted to induce assimilation because they represent an accidental gap in the otherwise regular distribution of OBLI clusters in French. Among the languages of the world, /dl/ and /tl/ do exist but seem to be infrequent and, perhaps, unstable combinations. In certain languages, for example, the liquid segment /l/ evolved into /r/ or into a glide[8]. Although a systematic documentation on clusters and their evolution in the various languages is not available, there exist anecdotal evidence that dental and back articulation plosive+/l/ clusters do not co-exist in certain languages (these two types alternating in the different dialects of a same family[9]). We may surmise that such clusters are intrinsically confusable and that this is why certain languages do not include /dl/ and /tl/ in their cluster repertoire. By this view, the perceptual assimilation we observed in modern French is just another facet of the same intrinsic confusability. Quite conceivably, then, the dental stop+/l/ clusters could produce perceptual assimilations in other languages than French.

We obtained evidence for perceptual assimilation in both the on-line and off-line processing of nonwords. In the open-response task, where subjects could use their phonological and/or orthographic knowledge post-perceptually, the effect was very strong and perhaps artificially inflated. The gating results confirmed that the phonetic quality of the cluster-initial consonants in the /tl/ and /dl/ items was dental rather than velar. Gating data also revealed the time-course of the perceptual assimilation phenomenon. It began to unfold after the /l/ segment was identified, and developed gradually thereafter. Finally, perceptual assimilation was found to be strongest in on-line processing as reflected by the phoneme

monitoring task, with a significant trend for velar targets to be detected slower in illegal rather than legal clusters.

A consistent finding across the different experiments was that /dl/ was more resistant to assimilation than /tl/. We may ask why. A first explanation appeals to some perceptual differences. The data on (American) English phoneme confusability under various signal-to-noise ratios (Miller and Nicely, 1955) show that English /d/ is perceptually more robust than /t/: 47% confusions for /d/ *vs.* 66% for /t/ (for S/N = -12dB). Interestingly, both dental plosives are misperceived as velars more often than as labials: 33% *vs.* 18% (computed from Miller and Nicely, 1955, Table XVIII). Of course, the French contrast /d/ **:** /t/ is different. Voiced plosives are prevoiced and voiceless plosives have a short voicing lag. Some cues to place of articulation may exist in the prevoiced portion. Even if this is not the case, the prevoiced portion by itself sounds like a nasal murmur and may be viewed as a weakly-resonant vowel-like sound, which, somehow, enhances the perceptibility of a /dl/ cluster[10].

A different explanation for the difference in perceptual assimilation between the two illegal clusters refers to their relative frequency of usage. Speaking of "frequency of usage" may seem strange in the case of illegal sequences, but both illegal sequences /dl/ and /tl/ are sometimes found utterance-initially, in such contracted forms as /dlo/ for "De l'eau" ([some] water) or /tlavy/ for "Tu l'as vu ?" (did you see him/it?). The former construction (partitive article+noun), however, is probably more common, and more productive than the second one[11]. That French listeners are more often confronted with utterance-initial /dl/ than with /tl/ sequences in running speech might partly explain that they are less likely to assimilate /dl/ than /tl/. We believe, however, that this can only be part of the story and is, perhaps, a consequence rather than a cause of the greater robustness of /dl/ in French, which may be largely explained by basic perceptual characteristics.

As was discussed earlier, the FLMP model could probably accomodate the contextual perceptual assimilation phenomenon, including its time course, with no explicit reference to levels of representation. (However, this approach implicitly requires that phonological, or more precisely, phonotactic knowledge be represented.) Contextual perceptual assimilation can also be understood within an interactive activation framework. Indeed, this is a situation

in which an initially strongly activated representation at the phoneme level is eventually overcome by a different Ñinitially weakly activatedÑ representation, whose activation growth is due to top-down feedback from representations at a higher level than the phoneme level. The TRACE model (McClelland and Elman, 1986) would explain the perceptual shift in terms of lexical-to-phoneme feedback, since it does not propose intermediary prelexical level(s) of representation. At this point, however, there is no reason to exclude the possibility of prelexical levels of perceptual integration other than phonemes. For instance, there is evidence in French that the syllable and its constituent parts may be involved in on-line speech processing (Mehler, Dommergues, Frauenfelder, & Segui, 1981). We only need to posit a higher level of representation (i.e., of perceptual integration) than the phoneme Ñwe propose syllable onsetÑ to explain that top-down activation may modify phoneme-level representations when they when they are not compatible with higher level representations. In this type of model, phonotactic knowledge can be represented either by specific levels of initial activation or by specific activation thresholds for such nodes as "onset nodes". TRACE, as well as similar models incorporating prelexical levels of representation, would predict a delay in the activation of phoneme-level representations when a contextual perceptual shift has occurred: Indeed, the top-down activation (causing the shift) must take time to overcome the initial bottom-up activation that could induce an unbiased perception at the phonetic level. This is consistent with the slower increase in velar responses in the /dl/ or /tl/ clusters as compared with that obtained in /kl/ or /gl/ sequences in the gating experiment and with the longer detection latencies found in Experiment 5 for velar detection. With TRACE, top-down feedback is lexical. In the case of the dental + /l/ items, even though no word is eventually recognized, all the words sharing a velar + /l/ onset are activated to some degree, and are the source of the feed-back activation of velar stops. With models incorporating, for example, onset level representations, the source of the feed-back is prelexical. In other words, the latter models leave open the possibility that phonotactic knowledge and constraints are not exclusively coded in the mental lexicon. A similar view is held by Pitt and Samuel (1995) who suggest that prelexical effects are able to "mimick" lexical effects.

The present results can also be viewed from a more phonological perspective. Normal phonetic perception of fully articulated speech items (as opposed to gated fragments) may take as elemental units syllable onsets rather than singleton phones. In other words, a syllable-initial cluster may be perceived as a whole. This is supported by the widespread uncertainty as to the single phoneme *versus* cluster status of certain phones with a complex articulation (see footnote 1). In this perspective, word-initial /tl/ may be treated as an atypical, or bad exemplar of the /kl/ category. Atypicality naturally suggests an intrinsic difficulty of perceptual processing. In the recent model of perceptual assimilation offered by Best (Best, 1990, 1994), the stress is on the difficulty in perceiving nonnative contrasts when one or both members are assimilated as bad exemplars of a native category ("Category Goodness" and "Single Category" contrasts). We suggest that the difficulty is basically due to that of accommodating for atypical exemplars. Therefore, the phonotactic violations that lead to assimilation of illegal clusters into their legal counterparts must introduce some additional processing difficulty.

This special kind of assimilation was especially strong in the speeded detection task where it produced a temporal processing cost. Illegal clusters may be viewed as distortions in the speech input that subjects did not notice most of the time, but that led to some difficulty of perceptual integration. This situation is reminiscent of the "subcategorical mismatch" cross-splicing manipulations used by different authors (Marslen-Wilson and Warren, 1994; Streeter and Nigro, 1979; Whalen, 1984, 1991). The distortions resulting from cross-splicing are not consciously perceived, but generally introduce perceptual integration difficulties, depending on the stimulus structure, the lexical status of the stimuli and its competitor environment. The patterns of RT costs observed provided new insights in lexical access and lexical representations (Marslen-Wilson and Warren, 1994). The case of assimilated illegal clusters is similar, in that a perturbation of the speech input goes unnoticed but nevertheless entails an increase in processing difficulty. The general implication is that perceivers are biased toward perceiving events and objects that fit with internal representations rather than with objective stimulation. The bias may be even stronger when the object of the perception is a meaningful event, as is the case for words as opposed to non-words. Substantial

mismatch between percept and stimulus can be tolerated, but at the expense of some additional processing. Experimental conditions such as subcategorical mismatch in which such accomodating processes are at work are thus of particular interest. Phonotactic mismatch is yet another, with the presumed advantage of being ecologically more relevant. Illegal clusters still are natural speech, while subcategorical manipulations result in "speech" items that may sound relatively natural, but that do not correspond to any human articulation. Contextual perceptual assimilation therefore has some potential methodological implications. Perceptually assimilated illegal clusters (inducing extra-processing cost) could serve as distortions of the speech input at the phonetic level, and help investigating further stages of speech perception: How would such distortions impede perceptual integration at various sublexical and lexical levels? Addressing such issues could shed light on the mechanisms whereby we perceive internalized plausible/meaningful events rather than their objective physical manifestations.

We will finally return to the issue of the origin of the contextual perceptual assimilation phenomenon. Lexical mediation as predicted by TRACE cannot be excluded, although the results of the gating experiments rather point to an assimilation process emerging at a sublexical level, within prelexical units such as CCV. The size of the effect, however, was smaller for the gated presentation of such syllabic units than for that of entire words. It can also be argued that lexical activation is possible even on the basis of partial information (see McClelland and Elman's, 1986, critique of Massaro and Cohen's, 1983, notion of phonological context effect: pp. 33-35). According to a different view, the implicit knowledge of phonotactic constraints may originate at a more basic level. Listeners may know that certain combinations of sounds are legal or not, not only by reference to their mental lexicon, but also based on a deeper intuition: "Speakers must have tacit knowledge about how phonemes are strung together in their language." (Pinker, 1994, p. 173). Such knowledge is more primitive, as it were, than lexical knowledge in the sense it developed from early infancy. The infant data are rather more consistent with this second view. Nine-month-olds, that is, infants with a small or no receptive lexicon (see Hallé and Boysson-Bardies, 1994), are sensitive to phonotactic violations (Jusczyk et al., 1993), and even to the

frequency of occurrence in the language of particular CVC sequences (Jusczyk et al., 1994), which probably correlates with the degree of familiarity that infants have with such sound combinations. Hence, before a lexicon has developed, young listeners are already "attuned to" the sounds and combinations of sounds heard in the ambient language. We surmise that this primitive (that is, proto-lexical) knowledge of phonotactic constraints may survive in adults and partly explain the often modified perception of illegal clusters such as /dl/ and /tl/ in French.

REFERENCES

Best, C. (1990). Adult perception of nonnative contrasts differing in assimilation to native categories. *Journal of the Acoustical Society of America*, **88**, S177(A).

Best, C. (1991). Phonetic influences on the perception of nonnative speech contrasts by 6-8 and 10-12 month olds. *Paper presented at the Society for Research in Child Development, Seattle*.

Best, C. (1994). The emergence of native-language phonological influences in infants: a perceptual assimilation model. In J.C. Goodman and H.C. Nusbaum (Eds.), *The Development of Speech Perception: The Transition from Speech Sounds to Spoken Words* (pp. 167-224). Cambridge, MA: MIT Press.

Best, C., McRoberts, G., Lafleur, R., and Silverisenstadt, J. (1996). Divergent developmental patterns for infants' perception of two nonnative consonant contrasts. *Infant Behavior and Development*, **18**, 339-350.

Best, C., McRoberts, G., and Sithole, N. (1988). Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance*, **4**, 45-60.

Blumstein, S.E., and Stevens, K.N. (1980). Perceptual invariance and onset spectra for stop consonants in different vowel environments. *Journal of the Acoustical Society of America*, **67**, 648-662.

Brown, R.W., and Hildum, D.C. (1956). Expectancy and the perception of syllables. *Language*, **32**, 411-419.

Cotton, S., and Grosjean, F. (1984). The gating paradigm: A comparison of successive and individual presentation formats. *Perception and Psychophysics*, **35**, 41-48.

Cutler, A., Mehler, J., Norris, D., and Segui, J. (1986). The syllable's differing role in the segmentation of French and English. *Journal of Memory and Language*, **25**, 385-400.

Cutler, A., Mehler, J., Norris, D., and Segui, J. (1992). The monolingual nature of speech segmentation by bilinguals. *Cognitive Psychology*, **24**, 381-410.

Dupoux, E., Kakehi, K., Hirose, Y., Pallier, C., Fitneva, S., and Mehler, J. (in press). Epenthetic vowels in Japanese: a perceptual illusion? *Journal of Memory and Language*.

Flege, J. (1986). The production and perception of foreign language speech sounds. In H. Winitz (ed.), *Human Communication and its Disorders* (pp. 224-401). Norwood, NJ: Ablex.

Flege, J. (1991). Perception and production: The relevance of phonetic input to L2 language learning. In C. Ferguson and T. Heubner (Eds.), *Crosscurrents in Second Language Acquisition and Linguistic Theories* (pp. 249-289). Philadelphia: John Benjamins.

Fowler, C.A, Best, C.T., and McRoberts, G.W. (1990). Young infants' perception of liquid coarticulatory influences on following stop consonants. *Perception and Psychophysics*, **48**, 559-570.

Frauenfelder, U., and Segui, J. (1989). Phoneme monitoring and lexical processing: Evidence for associative context effects. *Memory and Cognition*, **17**, 134-140.

Frauenfelder, U., Segui, J., and Dijkstra, T. (1990). Lexical effects in phoneme processing: Facilitatory or inhibitory?. *Journal of Experimental Psychology: Human Perception and Performance*, **16**, 77-91.

Grosjean, F. (1980). Spoken word-recognition processes and the gating paradigm. *Perception and Psychophysics*, **28**, 267-283.

Halle, M., Hughes, G.W., and Radley, J.-P.A. (1957). Acoustic properties of stop consonants. *Journal of the Acoustical Society of America*, **29**, 107-116.

Hallé, P.A., and Boysson-Bardies, B. (1994). The emergence of an early lexicon: Infants' recognition of words. *Infant Behavior and Development*, **17**, 119-129.

Imbs, P. (1971). *Études statistiques sur le vocabulaire français. Dictionnaire des fréquences. Vocabulaire littéraire des XIXe et XXe siècles* [Statistical studies on French vocabulary. Dictionary of frequencies. Literary vocabulary of the 19th and 20th centuries]. Centre de recherche pour un Trésor de la Langue Française (CNRS). Paris: Librairie Marcel Didier.

Jusczyk, P.W., Cutler, A., and Redanz, N. (1993). Preference for the predominant stress patterns of English words. *Child Development*, **64**, 675-687.

Jusczyk, P.W., Luce, P.A., and Charles-Luce, J. (1994). Infants's sensitivity to phonotactic patterns in the native language. *Journal of Memory and Language*, **33**, 630-645.

Jusczyk, P.W., Friederici, A.D., Wessels, J., Svenkerud, V.Y, & Jusczyk, A.M. (1993). InfantsÕ recognition of foreign versus native language words. *Journal of Memory and Language*, **32**, 402-420.

Kewley-Port, D. (1983). Time-varying features as correlates of place of articulation in stop consonants. *Journal of the Acoustical Society of America*, **73**, 322-335.

Kuhl, P., Williams, K., Lacerda, F., Stevens, K., and Lindblom, B. (1992). Linguistic experiences alter phonetic perception in infants by 6 months of age. *Science*, **255**, 606-608.

Lahiri, A., Gewirth, L., and Blumstein, S.E. (1984). A reconsideration of acoustic invariance for place of articulation in stop consonants: Evidence from a cross-language study. *Journal of the Acoustical Society of America*, **76**, 391-404.

Maddieson, I. (1984). *Patterns of Sounds*. Cambridge: Cambridge University Press.

Mann, V.A. (1980). Influence of preceding liquid on stop-consonant perception. *Perception and Psychophysics*, **28**, 407-412.

Mann, V.A. (1986). Distinguishing universal and language-dependent levels of speech perception: Evidence from Japanese listeners' perception of English "l" and "r". *Cognition*, **24**, 169-196.

Marslen-Wilson, W.D. (1984). Function and process in spoken-word recognition. In H. Bouma and D.G. Bouwhuis (Eds.), *Attention and Performance X: Control of Language Processes*, Hillsdale, NJ: Erlbaum.

Marslen-Wilson, W.D., and Warren, P. (1994). Levels of perceptual representations and process in lexical access: Words, phonemes, and features. *Psychological Review*, **101**, 653-675.

Massaro, D., and Cohen, M. (1983). Phonological context in speech perception. *Perception and Psychophysics*, **34**, 338-348.

Massaro, D., and Cohen, M. (1993). The paradigm and the fuzzy logical model of perception are alive and well. *Journal of Experimental Psychology: General*, **122**, 115-124.

Massaro, D., and Oden, G. (1980). Speech perception: A framework for research and theory. In N.J. Lass (Ed.), *Speech and Language: Advances in Basic Research and Practice* (Vol. 3, pp. 129-165). New York: Academic Press.

Massaro, D., and Oden, G. (1995). Independence of lexical context and phonological information in speech perception. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **21**, 1053-1064.

McClelland, J., and Elman, J. (1986). The TRACE model of speech perception. *Cognitive Psychology*, **18**, 1-86.

Mehler, J., Dommergues, J.Y., Frauenfelder, U., and Segui, J. (1981). The syllable's role in speech segmentation. *Journal of Verbal Learning and Verbal Behavior*, **20**, 298-305.

Messer, S. (1967). Implicit phonology in children. *Journal of Verbal Learning and Verbal Behavior*, **6**, 609-613.

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

Niederer, B. (1995). *Les langues Hmong-Mjen (Miao-Yao) : Problèmes de phonologie historique* [Hmong-Mjen (Miao-Yao) languages: Issues of diachronical phonology]. Unpublished Doctoral Dissertation, Paris III University.

Oden, G., and Massaro, D. (1978). Integration of featural information in speech perception. *Psychological Review*, **85**, 172-191.

Pinker, S. (1994). *The Language Instinct (How the Mind Creates Language).* NY: William Morrow and Company.

Pitt, M., and Samuel, A. (1995). XXXXXX

Polivanov, E. (1931). La perception des sons d'une langue étrangère. *Travaux du Cercle Linguistique de Prague*, **4** (partly reproduced in *Change*, 1969, **3**, 111-114).

Polka, L., and Werker, J.F. (1994). Developmental changes in perception of non-native vowel contrasts. *Journal of Experimental Psychology: Human Perception and Performance*, **20** 421-435.

Praamstra, P., and Levelt, W.J.M. (1994). Neurophysiological manifestations of phonological processing: Latency variation of a negative ERP¨component timelocked to phonological mismatch. *Journal of Cognitive Neuroscience*, **6**, 204-219.

Pullum G., and Ladusaw, W.(1986). *Phonetic Symbol Guide*. Chicago: The University of Chicago Press.

Stevens, K.N., and Blumstein, S.E. (1978). Invariant cues for place of articulation in stop consonants. *Journal of the Acoustical Society of America*, **64**, 1358-1368.

Streeter, L.A., and Nigro, G.N. (1979). The role of medial consonant transitions in word perception. *Journal of the Acoustical Society of America*, **65**, 1533-1541.

Troubeztkoy, N.S. (1939). Grundzüge der Phonologie [Principles of phonology]. *Travaux du Cercle Linguistique de Prague* 7.

Tubach, J.P., and Boè, L.J. (1990). *Un corpus de transcription phonétique (300,000 phones): Constitution et exploitation statistique* (Paris France Télécom Report No. 90 D 002). Paris: France Télécom.

Walley, A.C., Michela, V.L., and Wood, D.R. (1995). The gating paradigm: Effects of presentation format on spoken word recognition by children and adults. *Perception and Psychophysics*, **57**, 343-351.

Wang, F.S. (1979, September). *Miao-yu fangyan shengyunmu bijiao* [A comparative study of the onsets and rimes of Miao languages]. Manuscript circulated at the 10th Sino-Tibetan Conference, Paris.

Warren, P., and Marslen-Wilson, W.D. (1987). Continuous uptake of acoustic cues in spoken word recognition. *Perception and Psychophysics*, **41**, 262-275.

Werker, J., and Tees, R. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, **7**, 49-63.

Whalen, D.H. (1984). Subcategorical phonetic mismatches slow phonetic judgments. *Perception and Psychophysics*, **35**, 49-64.

Whalen, D.H. (1991). Subcategorical phonetic mismatches and lexical access. *Perception and Psychophysics*, **50**, 351-360.

Table 1. Preliminary experiment: Transcriptions of the /dl/ and /tl/ items.

item	% of transcriptions as:				
	/tl/	**/dl/**	**/kl/**	**/gl/**	**/pl/**
/tl/	9.7%	1.4%	81.9%	4.2%	2.8%
/dl/	-	15.3%	4.2%	80.6%	-

Table 2. Identification test 1: Percentages of dental-to-velar and dental-to-labial confusions according to stimulus type.

item-initial cluster		percentage of place confusions	
		dental-to-velar	dental-to-labial
(illegal)	/tl/-	63.5	1.0
	/dl/-	47.9	1.0
(legal)	/tr/-	8.3	1.0
	/dr/-	1.0	6.3

Table 3. Measured intervals (in ms) from the initial stop release burst to the onsets of the liquid /l/ or /r/, of the following vowel /a/ or /o/, of the following closure, and of the following release of /b/ or /p/, for the items used in the gating experiments.

item type	liquid	vowel	closure	release
dental + /l/	20	76	177	289
velar + /l/	24	71	157	272
dental + /r/	31	123	231	348

Table 4. Average percentages of dental judgments across all ten gates, according to test item.

item	type of presentation		
	successive	duration blocked	Mean
dlapot	90.7	91.2	90.9
dlapto	82.9	80.0	81.4
tlabod	79.3	80.0	79.7
tlabdo	80.4	83.8	82.1
tlobad	78.9	81.9	80.4
tlobda	68.9	75.0	71.9
dlopat	55.0	72.7	63.8
dlopta	52.9	76.9	64.9
Mean	73.63	80.19	76.89

FIGURE CAPTIONS

Figure 1. Short term spectra of the initial stop release burst (see note 6), the following /l/, and the onset of the following vowel in (a) 'tlabdo', (b) 'klabdo', and (c) [plA$] (from a filler item).

Figure 2. Dental-to-velar confusions and voicing confusions for /tl/ and /dl/ test items: four-fold forced choice identification test.

Figures 3.a to 3.d. Judgments of initial consonant place of articulation according to gate number: (a) /tl/ and /dl/ test items (successive format), (b) /kl/ and /gl/ control items (successive format), (c)/tr/ and /dr/ control items (successive format), (d) /tl/ and /dl/ test items (duration-blocked format).

Figure 4. Miss and false positive rates for dental and velar targets in dental+/l/ *versus* velar+/l/ items.

Figure 5. Detection latencies of velar targets in dental + /l/ items (f.a.) versus velar + /l/ (hit) items.

NOTE

* The authors gratefully acknowledge support from a grant from Axe de Recherche CNRS "Langage" (1994-1996), and from a F.N.R.S. grant 11-39553.93. We thank Carol Fowler, Arthur Samuel, Dominic Massaro, and an anonymous reviewer for their helpful comments on an earlier version of this paper. We are also grateful to Catherine T. Best, Laurent Sagart, and Louis-Jean Boë who provided phonetic and phonological data, and to Denis Lancelin who helped running the subjects in Paris.

Correspondence should be addressed to Pierre A. Hallé, CNRS-Paris V, Laboratoire de Psychologie Expérimentale, 28 rue Serpente, Paris, France.

[1] Certain nonnative single phonemes with a complex articulation can be assimilated to clusters of native phonemes. This is the case of the voiced *versus* unvoiced lateral fricatives of Zulu /Ÿ/ *versus* /K/, described by adult American listeners as sounding like clusters: /Jl/ or /Dl/ *versus* /Sl/ or /Ql/ (Best, 1990, 1991). In addition to the Zulu lateral fricatives, labio-velars such as [gb] or [gp], or prenasalized consonants such as [^{m}b], are possibly perceived as /g/+/b/ or /g/+/p/, or /m/+/b/ clusters. In fact, whether certain speech sounds are clusters or singleton phonemes may sometimes be arbitrary. For example, a /dl/ **:** /tl/ contrast has been posited in Tlingit (an Amerindian language) and described as a voicing contrast of "laterally released dental plosives" (Maddieson, 1984). In the IPA usage, the digram 'dl' denotes the same phone(s) as the American usage single letter '«\symbol»' (*lambda*), for a "voiced alveolar laterally released affricate" (Pullum and Ladusaw, 1986, p. 94).

[2] This finding has been recently reassessed (Dupoux, Kakehi, Hirose, Pallier, Fitneva, and Mehler, in press): In nonsense words such as /ebuzo/ (uttered by a Japanese speaker: [eb›zO]), Japanese subjects still hear an epenthetic vowel /u/ after the vowel and its acoustic correlates have been removed.

[3] The talker was not a trained phonetician but was judged as having a clear pronunciation in general. He practiced once the lists of non-words before recording.

[4] Dental-to-labial shifts were thus very marginal. This is not likely due to lexical effects: Both /pl/ and /kl/ are permissible word-initially, and /kl/-words are not more frequent than /pl/-words (see footnote 5). A plausible reason why dental-to-velar rather than dental-to-labial shifts were observed is the closer acoustic similarity of dentals to velars than to labials, both for burst spectrum properties and for consonant-vowel formant transitions (see Figure 1 for an illustration). Also, from an articulatory point of view, both dentals and velars are linguals whereas labials stand out, precisely, as labials.

[5] Before the vowels /a/ and /o/, the numbers of word-types with such plosive+liquid clusters in word-initial position are 227 (velar+/l/), 166 (labial+/l/), and 242 (dental+/r/); the corresponding number of tokens are respectively 27,066, 36,511, and 123,907 (source:

"Trésor de la Langue Française" [Imbs, 1971]). Hence, with regard to lexical frequencies, /gl/ and /kl/ are the least likely clusters that should be reported for the /dl/ and /tl/ items.

[6] Short term spectra were computed for 10 ms waveform portions beginning at 2 ms before the burst. In the case of /dl/ and /gl/, where the plosive was prevoiced, the voicing murmur was partly filtered out by high-pass filtering with a 500 Hz cut-off frequency. (This was done to make the measurements for prevoiced and voiceless plosives comparable.) Spectra were smoothed by means of an iterative cepstral smoothing technique, with a 2 ms cut-off quefrency. Spectral centers of gravity were computed from the smoothed spectra.

[7] The influence of the shortest stimuli on the level of analysis that subjects relied on was perhaps stronger in the duration-blocked format of presentation since, in this format, the shortest stimuli occurred first and the longer, syllable-like, stimuli occurred only in the last few blocks. In contrast, syllable-like stimuli appeared earlier, toward the end of each block in the successive format of presentation.

[8] For example in Polish, which has a /l/ **:** /Æ/ liquid contrast (still denoted by the written forms 'l' and 'Æ'). After /d/ and /t/, /Æ/ evolved to /w/: The written forms "dÆ" and "tÆ" are actually pronounced /dw/ and /tw/ in the standard variety of Polish; /dl/ and /tl/ did not undergo a similar diachronic change, but are rare in this language (!!!reference XXXXXX?????).

[9] Such is the case in the Miao-Yao group (Niederer, 1995). The various dialects of this group may have a series of dental stop+ /l/ clusters or a series of uvular stop+/l/ clusters, such as but not both. Wang (1979) has shown that these clusters all originate from velar/uvular clusters in proto-Miao. Certain words of proto-Miao with a back stop+/l/ cluster onset are borrowed from archaic Chinese words whose onset was most probably dental (e.g., N-lwat ([n^{d}lwat]) < dlwat < tuo). Hence, assimilation occurred at the time of the borrowing. In some modern Miao-Yao dialects, such as Shi Men, uvular changed back to dentals in stop+/l/ clusters, in others they did not, but in either case, uvular+/l/ and /dental+/l/ series are now mutually exclusive.

[10] Indeed, /tl/ and /dl/ are permissible in intervocalic position (that is, word-medially) as in "atlas", "atlantique", etc. (Examples for /dl/ are infrequent words or loan-words: "iodler" /jOdle/, "landlord" /lA$dlOrd/, or found in family names: "Adler" /adlEr/.) Word-medial schwa-deletion results in a somewhat larger number of word-medial /dl/ and /tl/ (as in "châtelain" /SAtlE$/, "godelureau" /gOdlyro/). Hence, /tl/ and /dl/ are probably "correctly" perceived in word-medial position, that is, after a vowel. If the prevoicing murmur plays a similar role as a faint vowel, this could explain why /dl/ is less often assimilated to /gl/ than /tl/ is to /kl/.

[11] Phonetic counts based on a corpus of spoken language (Tubach and Boë, 1991), including contracted forms, indicate that the /tl/ sequence is about 3-to-4 times more frequent than /dl/, regardless of word-position (raw counts: 134 *vs.* 39, out of about 300,000 diphones). On the other hand, lexical counts based on written language (excluding possible contractions in real speech), indicate that /tl/ is about ten times more frequent than /dl/ word-medially (Trésor de la Langue Française: Imbs, 1971). The discrepancy between spoken and written data (more /dl/ than expected in spoken language) therefore suggests that /dl/ is a more frequent contracted form than /tl/ in real speech. (Word-medially, /dl/ and /tl/ contracted forms have similar frequencies of usage.)