A voice for the voiceless: Production and perception of assimilated stops in French

Natalie D. Snoeren\textsuperscript{a}, Pierre A. Halle\textsuperscript{a,b}, Juan Segui\textsuperscript{a}

\textsuperscript{a}Laboratoire de Psychologie Expérimentale, 71 Avenue Edouard Vaillant, 92774 Boulogne-Billancourt, CNRS-Paris V, France
\textsuperscript{b}Laboratoire de Phonétique et Phonologie, CNRS-Paris III, France

Received 9 July 2004; received in revised form 24 May 2005; accepted 4 June 2005

Abstract

Previous studies, mainly conducted on English running speech, have reported that (place) assimilation between words is usually incomplete, and have suggested that the perceptual processing of word forms altered by that assimilation depends on the extent to which phonemes at word boundary have undergone assimilation. The present research provides an acoustic–phonetic description of regressive voice assimilation in French and proposes an objective measure of “voicing degree” for stops in word-final position. This measure is the relative duration of voicing within the stop closure. It is shown to closely correlate with perceptual judgments of voicedness. Across two experiments, we show that voice assimilation in French is graded and asymmetrical in that voiceless stops assimilate to a larger extent than voiced stops do. It also appears that assimilation is slightly modulated by lexical factors such as potential ambiguity and phonological neighborhood. If a word belongs to a minimal pair for final stop voicing (e.g., rate /rat/ ‘spleen’ minimally contrasts with rade /rad/ ‘harbor’), assimilation strength tends to be weaker. The same trend applies to a word challenged by ‘dangerous’ phonological neighbors that only differ from it by their word-final consonant.

\textcopyright 2005 Elsevier Ltd. All rights reserved.

\textit{E-mail address:} natalie.snoeren@univ-paris5.fr (N.D. Snoeren).

0095-4470/$ - see front matter \textcopyright 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.wocn.2005.06.001
1. Introduction

Over the last decade, there has been a renewed interest in the study of phonological alternation phenomena such as liaison and neutralization, among which figure reduction, deletion, and contextual assimilation. This class of phenomena has indeed been viewed as a privileged window on the basic interplay between speech production and perception. On the production side, it offers the potential to document the phonetic realization of alternative forms together with the factors that may condition or modulate alternations. On the perception side, it should help to understand the mechanisms that allow listeners to cope with alternations, i.e., to retrieve underlying speech forms in ‘altered’ surface forms. There is an apparent puzzle here: Efficient speech processing indeed requires both distinguishing words that minimally differ by a single feature on the one hand (e.g., run differs from rum) and dealing with lawful changes in the surface forms of words that neutralize those very distinctions on the other hand (e.g., run may change to rum in “run briskly”). A reasonable account of speech perception mechanisms should thus address these two conflicting requirements. One line of research is strongly linked with the assumption that assimilation is complete, that is to say, the assimilated form of “run” before a labial consonant is indistinguishable from an instance of “rum”. Consistent with this assumption, some researchers have proposed that the underlying form of assimilated segments is recovered, in the case of regressive assimilation, by means of regressive inference based on the following phonetic context (Coenen, Zwitserlood, & Bölte, 2001; Gaskell & Marslen-Wilson, 1996, 1998; Mitterer & Blomert, 2003). Gaskell and Marslen-Wilson (2001) further observe that listeners’ perception is largely based on the surface form whenever they face semantic ambiguity as in “A quick rum picks you up”, in which they would recognize “rum” rather than “run”. The studies that support the notion of regressive inference have in common that they systematically used speech stimuli pronounced with ‘deliberate’, complete assimilation. Another line of research holds that place assimilation is probably not often complete in naturally produced speech. This motivates the approach taken in the recent work of Gow (2001, 2002, 2003) to explain how listeners are usually not misled by assimilated forms. Gow observes that assimilation is often incomplete in natural situations of speech communication. As a consequence, the underlying place of articulation and the assimilation place are both present in the speech signal: preserved acoustic evidence, or ‘traces’ of the underlying place are exploited by listeners to recover the underlying form with little need or no need at all of the following phonetic context. (Coexistence of both underlying and assimilation places may also be exploited to anticipate the upcoming phonetic segments.) On this view, an accurate description of the acoustic consequences of assimilation in natural speech is crucial to understand how listeners cope with this kind of phonological alternation in real life.

In the phonological literature, assimilation (e.g., place assimilation) is usually described in categorical terms. However, although there is some debate on whether natural assimilation ever produces categorical change (Holst & Nolan, 1995; Nolan, Holst, & Kuhnert, 1996), a number of phonetic studies based on acoustic and/or articulatory measurements have reported incomplete assimilation (Barry, 1985; Ernestus, 2000; Gow, 2001, 2002, 2003; Gow & Hussami, 1999; Holst & Nolan, 1995; Kerswill, 1985; Nolan, 1992; Wright & Kerswill, 1989). More precisely there seems to be a continuum in ‘assimilation degree’, from complete assimilation (whose very existence may be questioned) to absence of assimilation, encompassing intermediate forms. In English, articulatory and acoustic evidence has been found for gradient changes in place of articulation
Nolan (1992; Wright & Kerswill, 1989). Nolan (1992) examined the acoustic and articulatory properties of (potentially) place-assimilated coronal stops. Electropalatographic data showed a continuum of tongue contact patterns. In sequences such as \textit{late calls}, the patterns varied from complete occlusion at the alveolar ridge with no dorsal activity (i.e., absence of assimilation) to posterior occlusion, indistinguishable from that observed in sequences such as \textit{make calls} (i.e., complete assimilation). In other words, place assimilation is graded, not discrete. Nolan (1992) further examined the perceptual consequences of incomplete and complete assimilation. Minimal coronal/non-coronal word pairs (e.g., road/rogue) were embedded in semantically neutral sentential contexts, with various degrees of assimilation of the coronal consonant. The sentences were presented to either experienced phoneticians or naive listeners, whose task was to identify the critical words. Whereas words with a non-coronal consonant (unaffected by the context) generated unambiguous responses, words with a (partially) assimilated coronal produced uncertainty. Participants nonetheless distinguished between words with underlying coronal and non-coronal consonants. This suggests that phonetic cues to the underlying coronal place in the assimilated consonants, which are indeed present in the speech signal, are detected and used by listeners, whether trained phoneticians or naive listeners.

Nolan’s (1992) study illustrates that appropriate acoustic–phonetic descriptions of phonologically motivated alternations in spoken forms provide a reasonable explanation of how they are perceptually processed: rather subtle cues in the assimilated segments are present in the speech signal and are exploited by listeners. Similar findings have been repeatedly reported with respect to the [voice] feature neutralization that occurs in ‘Final Devoicing’ in many languages (Dinnsen & Charles-Luce, 1984; Ernestus & Baayen, in press; Port & Crawford, 1989; Port & O’Dell, 1985; Slowiaczek & Dinnsen, 1985; Warner, Jongman, Sereno, & Kemps, 2004). Final Devoicing does not always result in complete neutralization, and listeners tend to take advantage of incomplete neutralization to recover underlying forms. As yet another illustration of incomplete neutralization, Manuel et al. (1992) found traces of the underlying form in severely reduced consonants and vowels in running speech. The phonetic and perceptual status of other phonological processes (e.g., epenthesis, deletion, mutation) have not received much attention and thus remains open to empirical investigations.

The purpose of the present study is to provide an acoustic description of naturally assimilated word-final consonants in running speech in French. In French, regressive voice assimilation can occur when two consonants of different voicing are in direct contact. For instance, the final /b/ of “robe” (‘dress’) would tend to devoice when followed by a voiceless consonant as in “robe sale” (‘dirty dress’). Rigault (1967) investigated the acoustic and physiological airflow consequences of voice assimilation for (French) /d/ in the context of /s/. He reported that /d/ assimilates to /t/ in this context. Moreover, according to Rigault’s data, /d/ in /d/+/s/ becomes physically and perceptually indistinguishable from /t/ in /t/+/s/, while radically differing from /d/ in /d/+/z/. Rigault (1967) thus claimed that voice assimilation in French is complete, both within words (e.g., \textit{médecin} [mitsɛ] ‘physician’) and between words (e.g., \textit{robe sale} [rɔpsal]).\footnote{Whether or not the pronunciation [mitsɛ] (or [mɛtsɻ]) of \textit{médecin} is a case of within-word voice assimilation is a matter of debate. From a synchronic point of view, it can be argued that the lexical form of “médecin” is simply stored as /mɛtsɻ/ and, as such, cannot and does not undergo assimilation. It would be a different story if the pronunciation of \textit{médecin} could alternate between [mɛtsɻ] and [medsɻ], as in, say casual vs. careful speech. But such is not the case.} Other linguists have
proposed less extreme views. For instance, Carton (1974) proposes that assimilation in French is more optional between words than within words. Grammont (1939) considers that a “fortis–lenis” distinction is maintained between plain voiceless and devoiced stops, or between plain voiced and assimilation-voiced ones. For example, the /b/ of “robe” would be realized as lenis [p], thus differing from /p/ in, e.g., “jupe” (“skirt”), in which /p/ is realized as fortis [p]. Given these discrepant views on voice assimilation in French, it seems necessary to conduct a phonetic study based on natural running speech. In the present study, we try to characterize voice assimilation at word boundaries (in French) in a quantitative way. We propose a specific acoustic index of ‘assimilation degree’ and explore the perceptual validity of this index.

It has been suggested that phonetic neutralization is modulated by high-level sources of information, such as communication intention or, even, morphologic and/or orthographic level of lexical representation (Ernestus & Baayen, in press; Port & Crawford, 1989). According to these authors, phonetic realization is closer to ‘canonic’ form, when there is potential confusion with some other word or some other meaning. Research by Port and Crawford (1989) showed that, under certain circumstances, Final Devoicing neutralization (in German) depends on the communicative context. In one condition, Port and Crawford asked participants to read a list of words in isolation. In the other condition, participants had to read sentences in which the neutralized contrast between two words was made explicit in a sentence such as Ich habe rat (rad) gesagt, nicht Rad (rat) (“I said ‘Rat (‘Rad’), not ‘Rad’ (‘Rat’)”), and they did so in front of an assistant who was writing down the test words produced. This was intended to elicit the speakers’ intention to make the words distinct. Indeed, this condition induced more incomplete neutralization than the first (word list condition). Thus, Final Devoicing was modulated by the communicative context: The need to avoid confusion between an intended word and a heterographic, though, in theory, homophonous word, induces productions with more preserved acoustic evidence of the underlying voicing (also see Gafos, in press, for a similar account). This is in line with, e.g., the previous research by Fowler and Housum (1987) which shows that speakers produce a word more carefully when this word appears for the first time in a conversation than when it has already been used once. The speaker’s intention of articulating new words more clearly or of making words that belong to minimal pairs more distinct (Port & Crawford, 1989) seems to influence speech production with respect to certain phonetic “details”. In a similar vein, Whalen (1991, 1992) found that lower frequency words are produced at a slower articulation rate than higher frequency words. That is, more uncertainty about a word’s identity seems to entail more careful articulation. Thus, whenever phonemic neutralization can take place, its actual implementation is modulated by higher-level factors, from lexical to pragmatic factors. Therefore, another purpose of the present study was to explore the potential role of some lexical factors on the phonetic realization of assimilated word forms. We limited ourselves to “potential ambiguity”

(footnote continued)
However, at the abstract morphophonemic level, médecin has a /d/ since the other members of its morphological family (médi­cal, médi­cament, etc.) all are pronounced with [d]; at a less abstract level, médecin has a /t/. Therefore, ‘within-word assimilation’ could be said to take place between levels of representation, by a transformation rule governing the alternation between [t] and [d] in the morphological family of “médecin”. Interestingly, ‘e’ in médecin is pronounced [e] more often than [ɛ], although it should be [ɛ] in the closed syllable /med/ or /med.s/. This can be thought of as symptomatic of the morphophonemic level of representation in which ‘e’ is indeed /ɛ/, as for médical, médicament, etc.
and phonological neighborhood. In some cases, assimilation may induce lexical ambiguity, as in the case of the form *run* in “a quick run/rum picks you up”. In the two production studies reported here, “potential ambiguity” was manipulated. A word may be said “potentially ambiguous” if it belongs to a minimal pair for final-stop voicing: e.g., “soute” /sut/ (‘hold’) contrasts with “soude” /sud/ (‘soda’); “frite” /frit/ (‘French fries’), on the other hand, is “unambiguous” because its voicing counterpart /frid/ is not a French word. In Experiment 2, we also manipulated phonological neighborhood restricted to the words differing with respect to the consonant of interest, the word-final consonant. For both “lexical ambiguity” and phonological neighborhood, we hypothesized that potential confusion with another word would reduce the degree or the likelihood of voice assimilation.

2. Experiment 1.a

2.1. Method

2.1.1. Materials

Two types of experimental words were used: potentially ambiguous and unambiguous. All were monosyllabic words with a stop in coda position. Each potentially ambiguous word minimally differs from a “fellow” word with respect to coda voicing. In other words, potentially ambiguous words come in pairs, such as (soute, soude) or (rade, rate), etc. Fourteen such minimal pairs, whose members were roughly matched in frequency (according to the Brülex database, Content, Mousty, & Radeau, 1990), were initially chosen, thus making 28 ambiguous words. Thirty-eight “unambiguous” words (such as “frite”, whose altered form /frid/ is not a French word) were chosen in the same frequency range. Each word, ambiguous or not, was inserted in two frame sentences, one possibly licensing voice assimilation, the other not. In the case of ambiguous words, both members of a pair where used in the same frame sentence, as in “la soute (soude) pue vraiment fort” (the hold (soda) smells really bad) vs. “la soute (soude) ne sent pas si mauvais que ça (the hold (soda) does not smell that bad)”. The sentences themselves were thus potentially ambiguous for these words. There were four sentences per ambiguous word pair, which we refer to as a “quadruplet” in the following. The initial selection comprised a total of 132 sentences (14 × 4 + 38 × 2). Twelve participants were asked to rate the plausibility of the sentences (with respect to syntactic and semantic well-formedness) on a 1–5 scale. Five quadruplets with potentially ambiguous words and 20 sentences with unambiguous words with ratings higher than 3.8 were eventually retained. These materials are listed in Appendix A. The potentially ambiguous words comprised five words with a voiced stop coda (e.g., soude) and their five voiceless counterparts (mean frequencies 27.3 and 46.8 per million, respectively, according to the ‘Lexique’

2Words ending with a fricative were not used because fricatives are generally less sensitive to contextual influence than stops. Robustness against change in fricatives has been documented for reduction in spontaneous speech and place assimilation (Jun, 1995; Kohler, 1990). It also generally holds for voice assimilation in French (Duez, 1995), except in a few cases such as “je” (“I”) followed by /s/ as in “je ne sais pas” (“I don’t know”) > [ʃsepa], [ʃepa] (Tranel, 1987, also see Duez, 2001, who reports cases of voice assimilation for fricatives).
database, New, Pallier, Ferrand, & Matos, 2001). Among the 10 unambiguous words retained (making up 20 sentences), seven had a voiceless stop coda and three had a voiced one. Their mean frequency was 27.5.

2.1.2. Speakers

Two men and two women took part in the production experiment. They were native speakers of French and had lived in the Paris region for at least 10 years. They all worked at René Descartes University.

2.1.3. Recording method

The four speakers read each sentence fluently at a normal speech rate. They first trained on the material once or twice before the recording was made. Participants were told they were allowed short pauses between sentences but not between words; they were not informed of the goal of the experiment. The recordings were made in a soundproof booth, using a Sennheiser microphone and a Tascam DAT recorder. The recorded speech was then transferred to computer files (20 kHz sampling rate, 16 bit resolution) for acoustic analyses.

2.1.4. Acoustic analyses

The most obvious acoustic cue for voicing in French, as in most languages, is the presence of vocal fold vibration (Lisker & Abramson, 1964) but a number of other acoustic cues, that may depend on the phonetic implementation of the [voice] feature in a particular language, have been proposed as well (Lisker & Abramson, 1971, for a review; also see Serniclaes, 1987). In English, voiceless syllable-initial stops (phonetically, voiceless aspirated) are associated with higher fundamental frequency (henceforth, f0) values at voicing onset than voiced stops (phonetically, voiceless unaspirated) (Ohde, 1984). Also, in English, vowels preceding stop consonants are typically longer when followed by a voiced consonant than when followed by a voiceless consonant (Peterson & Lehiste, 1960; Raphael, 1972; Umeda, 1975). The opposite pattern of durational contrast holds for stop closure durations. In French, formant transitions, burst intensity, f0 values at voice onset, and the preceding vowel duration have all been mentioned as acoustic properties that contribute to the voicing distinction (Duez, 1995; O'Shaugnessy, 1981; Saerens, Serniclaes, & Beeckmans, 1989; Wajskop, 1979). In this study, we made the following acoustic measurements for all the critical words: (1) f0 values in the vowel preceding stop closure; (2) duration of the vowel preceding the stop; (3) duration of the occlusion; and (4) voicing duration of the occlusion (i.e., duration of the voiced portion within occlusion; see Warner et al., 2004, for a similar measurement). In our measurements, the onset of occlusion is taken as the acoustic offset of the preceding vowel, signaled by an abrupt drop in amplitude and by the termination of the formant structure; the offset of occlusion is taken as the onset of release burst;

3The “Lexique” database was not available at the time of stimulus selection and we used the BRULEX database instead. Yet, to allow for comparison with the materials in the other experiments, we report here the frequencies as they appear in “Lexique”.

4In CV sequences, f0 contours are usually reported to start from higher f0 values for voiceless than for voiced obstruents but we are not aware of published data concerning VC sequences. Nevertheless, we conservatively left open the possibility that f0 patterns in VC sequences be symmetrical to those in CV sequences, be it for different articulatory and aerodynamic reasons.
finally, the offset of voicing within occlusion is set at the point where no more glottal pulse is discernible (see Fig. 1 for an illustration). In order to establish which of these acoustic measurements are most reliable to distinguish voiceless from voiced stop consonants in our data, their predictive power was compared for the stop consonants of those words that were embedded in non-assimilatory contexts.

Sometimes the word-final stop was not released when followed by another stop, particularly when the two stops were homorganic as in “sud de la France”, resulting in a long closure geminate stop. Absence of stop release occurred for 4% and 12% of the underlyingly voiceless and voiced stops, respectively. In those cases, closure duration was estimated by taking 50% of the total occlusion duration. Finally, an epenthetic vowel was produced after the stop in 3% of all productions.

2.2. Results and discussion

Table 1 summarizes the acoustic measurements performed on words in non-assimilatory context according to their potential ambiguity.

Because potentially ambiguous words come by minimal pairs with respect to coda consonant voicing, they allow for comparison of the acoustic measures in a strictly controlled phonetic context. Such is not the case for unambiguous words (see Appendix A). For the former subset of words, the measurements are in line with the known consequences of coda consonant voicing:

![Figure 1](image_url)

Fig. 1. A token of the French word *blette* [bEt] with incomplete assimilation. The dashed arrow stands for the voiced portion of the occlusion and the continuous arrow for the entire occlusion.
vowels are longer before voiced stops by about 12%, $t(19) = 1.88$, $p = .074$; closure durations are shorter before voiced stops by about 12% (a non-significant trend, $p = .16$). As for f0 at vowel offset, the weak tendency for lower f0s before voiced stops ($\sim 4\%$) is far from significant. These trends are also found for the subset of unambiguous words, except for vowel duration. Contrasting with these marginally significant or non-significant differences, a very robust voiced–voiceless difference of more than 200% is found for the closure voicing relative duration: in the phonetically controlled subset of ambiguous words, the difference is 261%, $t(19) = 10.5$, $p < .0001$; it amounts to 221% in the non-controlled subset, ($p < .0001$). This suggests that the relative duration of the voiced part of the occlusion provides a consistent index of voicing for words both in controlled and uncontrolled phonetic contexts. The acoustic durations involved in the computation of this index are shown in Fig. 1. In the following, we call ‘V-ratio’ the relative duration of the voiced part of the occlusion (following Nolan, 1992, it is converted here into a percentage).

As can be seen in Table 1, the mean V-ratio for voiceless stops in non-assimilatory contexts was 30%. That is, vocal fold vibration was continued well within the closure portion even for plain voiceless stops. There is a plausible explanation for the presence of vocal fold vibration in voiceless stops. The termination of the vocal fold vibration lags somewhat behind that of the vocal tract resonances relative to vowel. Therefore, the vocal folds may continue vibrating for a little while in the initial portion of occlusion. The duration of this “voicing lag” for voiceless stops was actually variable across items as well as across the four speakers we used. The word “trac” /træk/ was pronounced with a fully voiced [g] coda (V-ratio = 100%) by two speakers. With the exception of these two outlier instances, the V-ratio for voiceless stops in voiceless context ranged

<table>
<thead>
<tr>
<th></th>
<th>Vowel offset f0</th>
<th>Vowel duration</th>
<th>Closure duration</th>
<th>Voicing ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Ambiguous</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[−voice]</td>
<td>198 (67)</td>
<td>73 (26)</td>
<td>60 (17)</td>
<td>26 (22)</td>
</tr>
<tr>
<td>[+ voice]</td>
<td>191 (66)</td>
<td>82 (22)</td>
<td>53 (15)</td>
<td>95 (22)</td>
</tr>
<tr>
<td>$\Delta$ (%)</td>
<td>−4</td>
<td>+12</td>
<td>−12</td>
<td>+265</td>
</tr>
<tr>
<td><strong>Unambigious</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[−voice]</td>
<td>181 (71)</td>
<td>61 (18)</td>
<td>60 (16)</td>
<td>32 (25)</td>
</tr>
<tr>
<td>[+ voice]</td>
<td>162 (59)</td>
<td>40 (18)</td>
<td>54 (14)</td>
<td>100 (0)</td>
</tr>
<tr>
<td>$\Delta$ (%)</td>
<td>−11</td>
<td>−35</td>
<td>−10</td>
<td>+213</td>
</tr>
<tr>
<td><strong>All items</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[−voice]</td>
<td>188 (69)</td>
<td>66 (22)</td>
<td>60 (16)</td>
<td>30 (24)</td>
</tr>
<tr>
<td>[+ voice]</td>
<td>180 (64)</td>
<td>66 (29)</td>
<td>53 (14)</td>
<td>97 (18)</td>
</tr>
<tr>
<td>$\Delta$ (%)</td>
<td>−4</td>
<td>0</td>
<td>−12</td>
<td>+223</td>
</tr>
</tbody>
</table>

Note: The table is broken down for ambiguous and unambiguous items, produced by four speakers (SP1–4) in non-assimilatory contexts: (1) f0 value at vowel offset (Hz); (2) vowel duration (ms); (3) closure duration (ms); and (4) relative duration of voicing in the stop closure portion (in %). [± voice] refers to the word-final stop underlying voicing.
from 0% to 63% (SP1: [32–63%]; SP2: [0–41%]; SP3: [0–30%]; SP4: [31–51%]). Inter-speaker variability might partially be explained by individual differences in articulation style and rate, and might also reflect dialectal variations: although our four speakers had lived in the Paris region for many years, they were originally from different regions in France. As for voiced stops in voiced context, all were produced with a 100% V-ratio, with the exception of one instance of “bled” pronounced as [blɛt] (V-ratio = 0%).

The V-ratio can be thought of as a “voicing degree”. This leads to the straightforward derivation, from the V-ratio, of a “voice assimilation degree” (in percent): It is the percentage of voiced occlusion in underlyingly voiceless stops (‘100 × V-ratio’), or the percentage of voiceless occlusion in underlyingly voiced stops (‘100 × (1 – V-ratio)’). The analyses reported in the following are based on this proposed index of assimilation degree. Degree of assimilation, as defined above, was computed for the final stop of each critical word occurring in an assimilatory context. A large difference in assimilation degree was observed between ambiguous and unambiguous words, 25.2% and 77.8%, respectively, \(t(78) = 7.59, p < .00001\). However, the unambiguous stimuli were unbalanced with respect to coda voicing: There were more words with a voiceless coda than with a voiced coda in this subset. So the materials do not allow an assessment of the effect of lexical ambiguity independently of that of coda voicing. Fig. 2 shows the distribution of “assimilation degree” for a total of 80 occurrences, according to underlying voicing (12 voiceless and eight voiced words, produced by four speakers). Fig. 2 distinguishes five levels of assimilation degree: negligible assimilation [0–20%], three intermediate levels of assimilation [20–40%, [40–60%], and [60–80%], and near-complete to complete assimilation [80–100%].

The data reveal some important aspects of voice assimilation in French. Although the distribution looks rather bimodal with two peaks at the two endpoints of the V-ratio scale, a substantial percentage (39%) of the items fall in intermediate values of assimilation (in the [20–80%] interval). Apparently, word-final consonants are not always either completely assimilated or non-assimilated. Second, there is a striking asymmetry between voiceless and voiced stops: Overall, the mean degrees of assimilation are 67.7% and 27.3% for underlying voiceless and voiced word-final stops, respectively, \(t(78) = 4.96, p < .00001\). That is, voiceless stops
are, overall, more affected by assimilation than voiced stops (see Kohler, 1979, for a similar finding).5

The main findings reviewed above are based on the index of voicing degree we proposed, which provides a measure of assimilation degree for either voiced or voiceless stops. Although a number of alternative or complementary indices of voicing degree could be thought of, the relative voicing duration of stop closure was the most reliable acoustic index of “voicing degree” in our (French) data.6 Moreover, this index is able to capture voicing ‘degrees’ intermediate between fully voiced and fully voiceless, irrespective of phonetic context as the index was consistent across potentially ambiguous and unambiguous words. Now, is this index consistent with the phonetic judgment of French listeners? This question motivates a perceptual test focusing on the final stops in the materials we used. The strong asymmetry between voiceless and voiced stops further motivates such a perceptual test: Indeed, does this asymmetry truly reflect a ‘genuine’ asymmetry in voice assimilation directionality, or does it reflect an asymmetric bias in the measure of voice assimilation itself? If the asymmetry has a perceptual counterpart, assimilated segments that are underlyingly voiceless, should be perceived as voiced more often than voiced segments are perceived as voiceless, reflecting stronger assimilation in the former than in the latter case.

In the next experiment, we test the perceptual relevance of the index of assimilation used thus far, for both extreme and intermediate degrees of voice assimilation. The experimental words in the speech used in Experiment 1 were extracted from their sentential contexts and presented in isolation to French participants, who had to categorize the word-final consonant.

3. Experiment 1.b: phonetic categorization

3.1. Method

3.1.1. Materials and design

All the word-items produced in Experiment 1.a were used in Experiment 1.b. These were 10 ambiguous and 10 unambiguous words embedded in either assimilatory or non-assimilatory context, hence 40 stimulus-types (20 word-types × 2 contexts). Four speakers had produced the stimuli, hence a total of 160 stimulus tokens. Each word-stimulus was extracted from its sentence context. Care was taken to avoid any audible click at the onset or offset of the word. For words beginning with a voiceless stop, the starting point for extraction was chosen in the midst of the stop closure silence; in all the other cases (sonorants, fricatives, and voiced oral stops), it was

5In a few cases, voicing appeared to continue into the occlusion, ceases for a while, and then reappears prior to the end of the occlusion. In those cases, the second voicing interval was not taken into account into the quantitative analyses that led to a slight underestimation of the degree of voicing. If we had taken into account the latter interval of voicing, the asymmetry already observed between underlying voiceless and voiced stops would have been even more clear-cut.

6Closure duration and vowel duration have often proved to reliably cue voicing, in cases of incomplete [voice] neutralization, for languages such as Dutch or German (Ernestus & Baayen, in press; Jessen, 2001; Warner et al., 2004). In our data, this is found only in the phonetically controlled subset (the potentially ambiguous words), and even then, the trend somewhat lacks robustness. This is perhaps because there are intrinsic differences in the phonetic implementation of the [voice] feature in French as compared to those languages.
chosen as the point of maximum spectral stability either of the entire initial consonant, thus
entailing some truncation, or of the voiced murmur in stop closure. The endpoint for extraction
was in all cases the end of the release burst of the final stop (at the nearest zero crossing in the
speech signal). The extracted speech signal portion was further “feathered” by an 8-ms linear
attenuation ramp. Finally, the acoustic intensity was equalized across stimulus-types and across
speakers.

Two experimental lists of 80 stimuli were constructed. This was achieved for unambiguous
words by randomly assigning assimilatory and non-assimilatory sentential contexts to one list or
the other; for each potentially ambiguous word, the two versions of that word, from the
assimilatory and the non-assimilatory context sentence, were assigned to one list, and the two
versions of the voicing companion of that word were assigned to the other list. The organization
of the materials used in Experiment 1.b is illustrated in Table 2.

3.1.2. Participants

Twenty-two undergraduate or graduate students from the Psychology Department of René
Descartes University participated in the experiment. Seven were male and 15 female, with an
average age of 25 years. None of them reported any hearing or vision problem.

3.1.3. Procedure

Each participant was seated in front of a computer screen in a quiet room and first received
written and oral instructions. She or he was then presented with the stimuli at a comfortable
listening level via Sennheiser headphones. Participants’ task was to categorize the word-final
consonant of each item they heard, choosing one consonant within the set of eight consonants
proposed (/p, t, k, b, d, g, s, z/). At each trial, they also had to rate, on a 1–5 scale, how well the
item heard matched their response. Two training phases preceded the test phase. During the first
training phase, participants categorized the final consonant of unaltered words (produced in
isolation) and received feedback as to the correct response; they did not have to rate their
responses. During the second training phase, they were presented exclusively with non-words so as
to discourage the use of lexical knowledge to categorize item-final consonants; in this phase,

Table 2

<table>
<thead>
<tr>
<th>Word type</th>
<th>List 1</th>
<th>List 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Test</td>
<td>Context</td>
</tr>
<tr>
<td>Ambiguous</td>
<td>[−voice]</td>
<td>[−voice]</td>
</tr>
<tr>
<td></td>
<td>[−voice]</td>
<td>[+voice]</td>
</tr>
<tr>
<td></td>
<td>[+voice]</td>
<td>[+voice]</td>
</tr>
<tr>
<td></td>
<td>[+voice]</td>
<td>[−voice]</td>
</tr>
<tr>
<td>Unambiguous</td>
<td>[−voice]</td>
<td>[−voice]</td>
</tr>
<tr>
<td></td>
<td>[+voice]</td>
<td>[−voice]</td>
</tr>
</tbody>
</table>

Note: Test words appear in italics and context words in regular typeface. [±voice] refers to the voicing of the final stop in the test words (test) and to the voicing of the next consonant (context).
participants did not receive feedback but were trained to rate their responses on the 1–5 scale. After they had completed the experiment, which lasted about 20 min, participants filled in a language background questionnaire.

3.2. Results

In the following, we will call “congruent” the responses that are consistent with the underlying voicing of the target consonant, regardless of place of articulation. (Place of articulation, but not manner, was sometimes misperceived: the mean error rate for place was 10.8%.) The results can thus be summarized, for each of the 160 stimuli, by a percentage of “congruent responses”: the percentage of participants who reported the underlying voicing of the target consonant in that stimulus. The overall pattern of perceptual judgments according to underlying voicing, type of word (potentially ambiguous or not), and type of context (assimilatory or not) is presented in Table 3. The integration of subjective ratings produces “fit indexes” (Guion, Flege, Akahane-Yamada, & Pruitt, 2000) which may provide a more refined picture of participants’ perception than the raw percentages, uncorrected by ratings, reported here. However, because fit indexes essentially provided the same pattern of results as percentages, we limit ourselves to the analysis of the raw percentages.

We compared these percentages with the acoustic indexes of voice assimilation degree computed in Experiment 1.a. If this index has a perceptual relevance, we expect that a higher degree of assimilation would induce a lower percentage of “congruent responses”. In other words, a negative correlation between acoustic index of assimilation and perceptual recovery of underlying voicing should be observed, provided there is enough variation in assimilation degree for such a correlation to be meaningful. This latter criterion is met by the set of stimuli extracted from assimilatory context. In that set, the correlation between acoustic index of assimilation and percentage of “congruent responses” is negative and highly significant, \( r(78) = -0.83, p < .00001 \). This high level of correlation thus establishes the perceptual relevance of the acoustic index of voice assimilation degree, hence, the index of voicing as well, which was applied to final stop consonants in the present study.

Table 3
Experiment 1.b: Percentage of “congruent” responses (consistent with underlying voicing) according to context type (NAC: non-assimilatory context; AC: assimilatory context), word type (ambiguous or unambiguous), and underlying voicing

<table>
<thead>
<tr>
<th>Word type</th>
<th>Underlying voicing</th>
<th>Context type</th>
<th>No. of items</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>NAC</td>
<td>AC</td>
</tr>
<tr>
<td>Ambiguous</td>
<td>Voiceless</td>
<td>83.2</td>
<td>70.0</td>
</tr>
<tr>
<td></td>
<td>Voiced</td>
<td>93.5</td>
<td>90.9</td>
</tr>
<tr>
<td>Unambiguous</td>
<td>Voiceless</td>
<td>84.2</td>
<td>14.3</td>
</tr>
<tr>
<td></td>
<td>Voiced</td>
<td>76.4</td>
<td>62.9</td>
</tr>
<tr>
<td>Weighted means</td>
<td></td>
<td>85.1</td>
<td>54.7</td>
</tr>
</tbody>
</table>
As can be seen in Table 3, the overall pattern of perceptual judgments is consistent with the fact that perceptual judgments negatively correlate with the acoustic index of voice assimilation degree, and thus closely reflect—in reverse—the acoustic measurements reported in Experiment 1.a. Unsurprisingly, the percentage of congruent responses is consistently higher for non-assimilatory than for assimilatory context. We only analyze here the assimilation data of interest: the data obtained in the assimilatory context. Ambiguous words are associated with higher percentages of congruent responses than unambiguous words (80.5% vs. 28.9%, \( t(78) = 7.02, p < .00001 \)) and underlyingly voiced stops yield more congruent responses than underlyingly voiceless stops (80.4% vs. 37.5%, \( t(78) = 5.20, p < .00001 \)), consistent with the lower assimilation degree noted in Experiment 1.a for underlyingly voiced than voiceless stops (27.3% vs. 67.7%). The lowest rate of congruent responses was found for voiceless stops in unambiguous words in assimilatory context (14.1%).

The phonetic categorization data thus closely parallels the acoustic data for the asymmetry in terms of assimilation strength between voiceless and voiced stops. This suggests that the index of voicing accounts for a large part for the perception of voicing.

3.3. Discussion of Experiments 1.a–b

The acoustic and perceptual data of, respectively, Experiments 1.a and 1.b concur to suggest that relative voicing duration of stop closure provides a quite robust measure of voicing for French stops in non-initial position and is perceptually relevant. In the materials we used, other acoustic cues to voicing, such as vocalic or closure durations, did not reliably measure phonetic voicing. This is, perhaps, because a large number of both items and subjects are necessary for these cues to diagnose phonetic differences in voicing, as suggested by the recent work of Warner et al. (2004) in Dutch. This might also be due to the fact that the acoustic–phonetic basis of the [voice] feature in French is mainly the presence/absence of vocal fold vibration, and that the other cues, such as vowel, closure, or burst durations are secondary.

The acoustic measurements in Experiment 1.a provide evidence that voice assimilation in French is not always complete. Rather, it is a graded phenomenon: There are intermediate degrees of voicing in assimilated segments. Such empirical evidence challenges the phonetic and/or phonemic accounts of assimilation as an all-or-none phenomenon (Casagrande, 1984; Rigault, 1967). It is, however, quite in line with the growing evidence for the graded nature of place assimilation in English (Gow, 2002, 2003; Nolan, 1992).

A second result, robust although unexpected, emerged from the assimilation data: Word-final voiceless stops, such as /t/ in frite, are more frequently and more completely assimilated than word-final voiced stops, such as /b/ in tube. This difference is closely paralleled by the perceptual data. However, the materials of Experiment 1 were not balanced with respect to word-final voicing (seven voiceless against three voiced word types in the unambiguous word subset), except—by construction—for ambiguous words. Moreover, unambiguous words were also not balanced for place of articulation of the final stop, and for the type of the preceding vowel, which both may affect the duration of vocal fold vibration (see Ohala, 1983). Also, the assimilatory context (i.e., the following consonant) was not controlled, and comprised a mixture of stops, fricatives, and nasals. We therefore conducted an additional experiment in which all these aspects were better controlled (Snoeren & Segui, 2003). The speech materials were strictly balanced with
respect to stop voicing, and, as well as possible, to place. They were also controlled for the following consonant. In that experiment, we had 10 native speakers of Parisian French produce 40 unambiguous monosyllabic target words (similar to the unambiguous words in Experiment 1). Words were always embedded in an assimilatory context for voicing. The manipulated variables were the (underlying) voicing of target word-final stops, and the manner of articulation (stop vs. fricative) of the following consonant. Half of the following consonants were fricatives (/f, v, s, j/), and the other half were the stops (/p, t, k, b, d, g/). The nasal /n/ was avoided as it can induce nasalization rather than voice assimilation (Duez, 2001). We found equivalent degrees of assimilation for the fricative and stop contexts: on average 51.2% and 53.4%, respectively (voiced and voiceless stops pooled). Again, more assimilation was found for voiceless than for voiced stops: 75% and 29%, respectively (p < .001), just as in Experiment 1.a. These additional data thus confirm the asymmetric assimilation pattern between voiceless and voiced words for speech materials that are strictly balanced with respect to word-final stop voicing.

A third result concerns the large difference in assimilation degree between ambiguous and unambiguous words (25.2% vs. 77.8%). A tempting interpretation of this difference is that speakers tend to preserve the original voicing whenever a voicing shift could induce lexical confusion. Indeed, this would be a clear-cut illustration of an adaptation of production to anticipated perception (Lindblom, 1990). However, this asymmetry in assimilation degree between ambiguous and unambiguous words might probably be induced by the experimental design for at least two reasons. First, whereas the ambiguous word subset was balanced in terms of underlying voicing, the unambiguous word subset was not: It had more voiceless than voiced stops, and this could explain part of the difference between the two subsets since voiceless stops, as we have seen, assimilate more than voiced stops. Second, in Experiment 1.a, participants first practiced once or twice on the materials before the recording was made; moreover, they read quadruplets of sentences consisting of the two members of a minimal pair (e.g., soude and soute) embedded in the same two sentences. Although care was taken to set apart as much as possible within the recording list any two sentences differing only by their embedded word, it might be the case that speakers had been aware of the presence of lexically ambiguous items. This awareness could have induced more intentionally careful articulation with ambiguous than with unambiguous words, hence less assimilation with the former than with the latter type of words. Before we can discuss this aspect, we need to run a further production experiment, in which a more controlled recording procedure will be used in order to assess the influence of potential ambiguity on the extent of produced assimilation.

The possibility that speakers’ production be modulated by the implicit knowledge of perception is certainly worth a more thorough test: this is the main motivation for Experiment 2. Production should resist against contextual variation if variation potentially induces perceptual confusion. This is typically the case for those words that are ‘potentially ambiguous’ with respect to the [voice] feature of the final stop: soute can indeed be confounded with soude if the final stop gets voiced. Similarly, production should conceivably resist any type of variation, when the target word to be produced is intrinsically ‘difficult’ to recognize. This is typically the case for words that have many lexical competitors or high-frequency lexical competitors. In Experiment 2, we look at the possible influence of lexical neighborhood on the production of word-final stops. The competitors of interest are those words, which only differ from the target word with respect to their word-final consonant, not all the words obtained after a single phonemic change (Luce &
Pisoni, 1998). For instance, the word jupe ([ʒyp] ‘skirt’) has three competitors in that restricted sense of lexical neighborhood: juge (‘judge’), jure (‘I swear’), and jute (‘jute’); jupe is otherwise “unambiguous” regarding the voicing of its final stop because jube is not a French word. Looking at such words as jupe therefore allows for testing the role of lexical competition without a possible confound with final stop voicing ambiguity. It has previously been shown that phonological neighborhood density (the number of neighbors—NNB) can affect spoken word recognition. For instance, Luce and Pisoni (1998) used a perceptual-identification task in which listeners had to identify words in noise. Words with sparse neighborhoods (few neighbors) were identified more accurately than words with dense neighborhoods (many neighbors). If lexical neighborhood influences at all production in a way that is consistent with its influence on word recognition, we expect that words with a dense phonological neighborhood should be articulated more carefully to compensate for the increased difficulty in recognition. Thus, we should observe a lesser degree of assimilation of word-final stops for words with many competitors than for words with few competitors.

4. Experiment 2.a

Experiment 2 was designed to examine more in depth the findings of Experiment 1, some of which had not been anticipated. A new set of speech materials, similar to that used in Experiment 1, was constructed so that the factors of interest be as strictly controlled as possible. First, an equal number of voiced and voiceless word-final stops were included in the materials for unambiguous words in order to reexamine the asymmetric assimilation pattern between voiced and voiceless stops found in Experiment 1. A second, essential concern was the role of lexical ambiguity on assimilation degree. We wanted to limit as much as possible speakers’ awareness of the presence of minimal word pairs in the recording list. In Experiment 2, speakers practiced on training sentences different from the test sentences and containing different target words. Solely assimilatory contexts were used so that any given target word appeared only once in the recording list. Finally, half of the test sentences were filler sentences. This design was intended to greatly reduce the likelihood that speakers notice the presence of minimal pairs such as (soute, soude). (In addition, the members of such pairs were kept apart as far as possible in the recording list.) A further improvement on Experiment 1 was to carefully select speakers from a homogeneous population with respect to age and regional “dialect”.

Experiment 2 was designed to examine the role of two lexical factors possibly influencing carefuless of articulation toward avoiding lexical confusion: ‘potential ambiguity’ (as in Experiment 1) and lexical neighborhood density, both based on the final consonant of target words; these two factors were orthogonally manipulated.

4.1. Method

4.1.1. Materials

As in Experiment 1, all the target words were monosyllabic and ended with a single-stop coda. Eighteen minimal word pairs (with respect to final stop voicing) were initially selected. They formed 36 potentially ambiguous words that were roughly matched in frequency (5.21 and 8.74
for words ending with a voiceless and voiced stop, respectively). Because subjective frequency is often considered as a better predictor of lexical access than is objective frequency based on printed corpuses (Gordon, 1985), 15 subjects were asked to rate the ‘frequency’ of the 36 potentially ambiguous words plus 36 filler words on a 1–5 scale. These subjects, students of the first grade in Psychology, did not participate in the other experiments. Twelve word pairs whose members were close in subjective frequency were retained. The mean frequency ratings were 3.11 and 3.06 for the words with a voiceless and a voiced word-final stop, respectively (a non-significant difference, \( t(22) = .57, \text{n.s.} \)). The corresponding objective frequencies (according to the ‘Lexique’ database) were 6.7 and 7.2, respectively. Twenty-four unambiguous words were selected in the same frequency range (mean frequency 8.43). The 48 selected test words are listed in Appendix B. Half of them are potentially ambiguous and the other half are unambiguous, as in Experiment 1. A second lexical factor, phonological neighborhood, has been manipulated orthogonally so that three levels of neighborhood density are crossed with lexical ambiguity and final stop voicing. Although phonological neighborhood can be defined in a number of ways (Bailey & Hahn, 2001; Luce & Pisoni, 1998; Luce, Pisoni, & Goldinger, 1990), some of them taking into account experimentally derived phoneme confusability (Luce et al., 1990), we limited ourselves to the rather crude approximation of a word neighborhood as the set of words derived from that word by a single phoneme substitution. Moreover, because we test the influence of neighborhood on the production of the final consonant of target words, the relevant definition of word neighborhood is here the set of competitors that differ from that word by the word-final consonant only. As an example, the neighborhood of *jupe* [ʒyp] does not include *jappe* [ʒap] (‘yap’) which differs from *jupe* by the vowel, not by the coda. Careful articulation of /p/, thus would not help avoiding a possible confusion between *jupe* and *jappe*. Using this definition of restricted neighborhood, neighborhood density was looked up in the ‘Vocolex’ database (Dufour, Peereman, Pallier, & Radeau, 2002). The 48 test words in the materials were evenly distributed across three levels of neighborhood density: low (2–7 neighbors), medium (8–11 neighbors), and high (12–17 neighbors). Neighborhood density was crossed with both lexical ambiguity (two levels) and final stop voicing (two levels) so that there was an equal number of test words in each factorial cell. The factorial design is illustrated in Table 4.

Forty-eight monosyllabic filler words, matched in frequency and phonological neighborhood density with the test words, were used. Both the filler and test words were inserted in either “J’ai dit…vingt fois” (‘I said…twenty times’) or “J’ai dit…cent fois” (‘I said…hundred times’). Test words always appeared within an assimilatory context: their final stop, if voiced, was followed by

**Table 4**

Examples (in italics) of the different types of target words used in Experiment 2.a

<table>
<thead>
<tr>
<th>Word type</th>
<th>Voicing</th>
<th>Neighborhood density</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>[2–7]</td>
</tr>
<tr>
<td>Ambiguous</td>
<td>Voiceless</td>
<td><em>gratte</em></td>
</tr>
<tr>
<td></td>
<td>Voiced</td>
<td><em>grade</em></td>
</tr>
<tr>
<td>Unambiguous</td>
<td>Voiceless</td>
<td><em>floc</em></td>
</tr>
<tr>
<td></td>
<td>Voiced</td>
<td><em>vogue</em></td>
</tr>
</tbody>
</table>
the voiceless fricative /s/ of cent, or, if voiceless, by the voiced fricative /v/ of vingt. The choice of a semantically neutral (and monotonous) sentential context was intended to still reduce the likelihood that participants notice the presence of minimal word pairs and adopt item-dependent production strategies. One possible drawback of this option was that assimilation had to take place across syntactic phrases: it might be much weaker in that environment than within a syntactic phrase. In order to avoid the blocking of assimilation, speakers were specifically instructed to produce the sentences very fluently, with the words tightly “stringed” together.

4.1.2. Speakers

Eight native speakers of French (four male and four female students, mean age: 25 years) took part in our experiment. All the speakers were students at René Descartes University and spoke the Parisian or Ile-de-France regional variety of French.

4.1.3. Procedure

Speakers were given written instructions saying they would have to read a number of test sentences in French. They were not aware of the experiment’s goal. It was stressed that they had to pronounce the sentences at a normal speech rate but as fluently as possible, with no pausing between words. Four sentences pronounced by a native speaker of French were presented to the speakers to make them understand how fast and fluently they had to pronounce the sentences. The speakers first practiced on 10 training sentences, for which they were given feedback by the experimenter as to the adequacy of their articulation rate and fluency. Speakers did not practice on the test sentences: They began to read and record them right away after the training was completed. The recording was made with the same apparatus as in Experiment 1.a. The recorded speech was digitized (44.1 kHz sampling rate, 16 bit resolution) and transferred to computer files for further acoustic analyses. After each speaker was done with the recording, he/she was asked by the experimenter whether he/she had noticed something peculiar about the test sentences. The speakers indeed noted that the frame sentences were always the same throughout the list, but none of them had been aware of the presence of minimal pairs such as (soute, soude).

4.2. Results

For each item, the relative duration of the voiced part of the occlusion in the final stop was measured. This index of voicing—the ‘V-ratio’—was converted into assimilation degree, as defined in Experiment 1.a. The results confirm the graded nature of assimilation found in Experiment 1.a. This is clearly suggested by the substantial number of occurrences of intermediate assimilation degrees (44.5% occurrences lie in the [20–80%] range of assimilation degrees), as is seen in the distribution of assimilation degrees for either ambiguous or unambiguous items (Fig. 3). Moreover, there was as much assimilation overall in this experiment as there was in Experiment 1.a (52.2% and 54.7%, respectively: see Tables 3 and 5). Thus, the possible blocking of assimilation at a phrase boundary was successfully avoided by the emphasized instruction to produce sentences as fluently as possible.

Table 5 shows the assimilation degree data according to the factorial design indicated in Table 4.
It has been suggested that assimilation could depend on place of articulation (e.g., Umeda, 1977). We therefore analyzed assimilation degree with respect to place of articulation as well as to underlying voicing. This is shown in Table 6: As can be seen, the most noticeable trend is that assimilation degree was consistently higher for voiceless stops, whatever the place of articulation.

Analyses of variance (by item and by speaker) were carried out on the data with Neighborhood (small, medium, and large NNB), Ambiguity (potentially ambiguous vs. unambiguous items), and Voicing (underlyingly voiced vs. voiceless stops) as within-subject factors. Neither Neighborhood nor Ambiguity had a significant effect. The effect of Voicing was highly significant, $F_1(1, 7) = 32.55$, $p < .001$, $F_2(1, 36) = 172.60$, $p < .0001$, reflecting a greater assimilation degree for voiceless than for voiced stops. Voicing significantly interacted with Ambiguity in the subjects analysis, $F_1(1, 7) = 6.94$, $p < .05$, not in the items analysis, $F_2(1, 36) = 1.39$. This interaction reflects a trend toward a somewhat stronger voiced–voiceless asymmetry for unambiguous than for ambiguous items, although the effect of Voicing was significant for both ambiguous and

Table 5
Assimilation degree (%), according to neighborhood density, underlying voicing, and lexical ambiguity (Experiment 2.a)

<table>
<thead>
<tr>
<th>Word type</th>
<th>Voicing</th>
<th>Neighborhood density</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ambiguous</td>
<td>Voiceless</td>
<td>75.0</td>
</tr>
<tr>
<td></td>
<td>Voiced</td>
<td>23.5</td>
</tr>
<tr>
<td></td>
<td>Means</td>
<td>49.3</td>
</tr>
<tr>
<td>Unambiguous</td>
<td>Voiceless</td>
<td>82.2</td>
</tr>
<tr>
<td></td>
<td>Voiced</td>
<td>30.7</td>
</tr>
<tr>
<td></td>
<td>Means</td>
<td>56.5</td>
</tr>
</tbody>
</table>

Note: Each data cell stands for 32 values (4 items × 8 speakers).

Fig. 3. Distribution of assimilation degree for word-final stops (Experiment 2.a) in ambiguous (black bars) and unambiguous words (white bars). The distribution is computed over 384 items (48 words × 8 speakers).
The results suggest that there is a slight trend toward less assimilation for ambiguous items: 50% and 54% mean assimilation degrees for ambiguous and unambiguous items, respectively. A closer look at the data in Fig. 4 reveals subtle differences in assimilation degree between the two types of items. The differences are confined at the two ends of the distribution of assimilation degree, the [0–20%] and [80–100%] intervals, which correspond to little or no assimilation and to almost complete assimilation, respectively. More ambiguous than unambiguous items underwent little or no assimilation: 28% vs. 20%, \( t(7) = 2.50, p < .05 \). Symmetrically, more unambiguous than ambiguous items underwent complete or almost complete assimilation: 34% vs. 29%. However, this latter difference failed to reach statistical significance. These trends indicate that ambiguous items induce a greater resistance to assimilation than unambiguous items, although the difference is indeed much weaker than that found in Experiment 1.

With the measure of neighborhood we used, neighborhood density (or NNB), no neighborhood effect emerged. We therefore turned to another measure of lexical neighborhood, which might better reflect the efficacy of a word’s neighbor(s) at hampering the perception of that word. This
measure is defined, for a given word, as the NNB which are more frequent than it is: we call it ‘number of dangerous neighbors’ (NDN). The measures of neighborhood that only consider such “more frequent neighbors” (their number, cumulated frequency, or, even, the mere existence of a clearly more frequent neighbor) have often proved to be at least as informative on neighborhood ‘dangerousness’ as the crude density measure (Grainger, O’Regan, Jacobs, & Segui, 1989; Luce, 1986; Meunier & Segui, 1999). Using this measure, we reexamined the neighborhood effect by performing correlation analyses. For the set of voiceless items, which exhibited the largest range of variation in assimilation degree, NDN correlated negatively with assimilation degree, \( r(22) = -0.51, p < .05 \). To sum up, we found indication that lexical ambiguity but also neighborhood competition (provided an appropriate measure, NDN, is used) tend to inhibit assimilation, conceivably because they tend to induce an increased care in articulation.

In the next experiment, we seek to confirm the perceptual relevance of the acoustic index of assimilation we used thus far, like we did in Experiment 1.b. Here, we decided to use the speech of a single speaker and select, within his productions, a set that exhibited the widest spectrum of assimilation degrees as possible, so that the comparison between acoustic and perceptual data be not confined within narrow intervals of assimilation degree. As in Experiment 1.b, participants categorized word-final consonants in test words extracted from their frame sentence.

5. Experiment 2.b: phonetic categorization

5.1. Method

5.1.1. Materials

The speech produced by one of the eight speakers used in Experiment 2.a, S1, was chosen because it was judged as well articulated, and it was representative of the speakers’ speech in terms of voice assimilation: The mean assimilation degree for S1 was 49.8%, the closest to the group’s mean (52.3%); the percentage of extreme patterns of assimilation (absence of assimilation, 0%, or complete assimilation, 100%) was 52.1% for S1, closest, again, to the group’s mean (54.7%). Among the 48 test words that had been produced by S1, 24 were selected so as to cover the full range of assimilation degrees from 0% to 100%. In effect, two-thirds of the items retained had an assimilation degree falling within the [20–80%] intermediate range and one-third had extreme values of assimilation degree. Twelve of these test items were ambiguous (mean assimilation degree 45.4%) and the other 12 were unambiguous (mean assimilation degree 49.3%), with an equally spread spectrum of assimilation degrees for both. Twelve filler items were also included. The test phase was preceded by a training phase comprising 10 items. All these 46 items were extracted from their embedding frame sentence following the same procedure as described for Experiment 1.b.

5.1.2. Participants

Twenty undergraduate and graduate students (two men and 18 women) at the Psychology Department of René Descartes University took part in the experiment. Their age ranged between 18 and 28 years (mean: 24 years). All the participants were native speakers of French and none of them reported any hearing difficulty. None of them had participated in the other experiments or additional tests reported in this study.
5.1.3. Procedure

Participants were tested individually in a quiet room. They were presented with the stimuli at a comfortable listening level, using Sennheiser headphones. The experiment was set up using the Praat software package (Boersma & Weenink, 1992–2004). Participants were instructed that they would receive speech stimuli ending with a consonant and had to decide, for each stimulus, which final consonant (among the /p, t, k, b, d, g, s, z/ forced-choice set) they believed they heard. Participants gave their response by clicking with a mouse on one of the eight consonants in the forced-choice set, which was displayed on the computer screen. For each stimulus, they also had to rate on a 1–5 scale how well their response matched the speech they had heard. The test phase consisted of two blocks, each containing the same 24 test items and the 12 filler items but randomized in a different order. Each test item was thus presented twice. Ten monosyllabic words were presented during the training phase. After the experiment, participants filled in a language background questionnaire. The whole experiment took about 15 min.

5.2. Results

Responses are counted as “congruent” when they are consistent with the underlying voicing of the target consonant, regardless of place of articulation. As was the case in Experiment 1.b, place of articulation was sometimes misperceived, the mean error rate for place was 5.8%. Also as in Experiment 1.b, we find that assimilation degree correlates negatively with percentage of “congruent” responses, \( r(22) = -0.73, p < 0.0001 \), consistent with the expectation that a higher degree of assimilation induces a lower rate of recovery of underlying voicing. The correlation is illustrated in Fig. 4.

5.3. Discussion of Experiments 2.a–b

The results of Experiment 2 confirm the main findings of Experiment 1. First, voice assimilation in French is graded, not all-or-none. Second, voice assimilation is asymmetric in the sense that originally voiceless stops assimilate more than voiced stops do. Finally, the strong correlation between the production and perception data found in Experiment 1.b is also found in Experiment 2.b. Therefore, it seems safe to conclude that the acoustic index used throughout this study is a reliable measure of voice assimilation in French.

In Experiment 2, we also tried to assess in a more controlled way the possible influence of lexical factors on the assimilation of word-final stops. Contrary to Experiment 1, we only found subtle differences between ambiguous and unambiguous items in the distribution of assimilation degrees, suggesting that potentially ambiguous words are more resistant to assimilation than unambiguous items. Phonological neighborhood did affect the extent of voice assimilation, at least for voiceless items. This could be seen when using the NDN measure of neighborhood (number of ‘dangerous’ neighbors), not the rather crude and less sensitive NNB measure. Speakers’ production of a target word is influenced by the existence of ‘dangerous’ phonological neighbors (more frequent than the target word) so that they tend to avoid assimilation. Similar measures of neighborhood, which we might collectively call “neighborhood distribution index”, have often been able to disclose lexical competition or morphological effects in both the auditory domain (Luce, 1986; Meunier & Segui, 1999) and the visual word recognition domain (Grainger et al., 1989).
6. General discussion

This study addressed the issue of between-word regressive voice assimilation of stops in French. We presented an acoustic–phonetic description of assimilation and examined the possible influence of lexical confusability on the production of assimilated stops. Acoustic measurements showed that the relative duration of the voiced part of stop closure was a reliable signature of voicing that could distinguish between word-final voiced and voiceless stops. From this measure of voicing, a measure of voice assimilation, independent of underlying voicing, was readily derived. The perceptual relevance of this measure as an index of assimilation was assessed in a phonetic categorization experiment in which subjects categorized word-final stops in words extracted from their sentential context. A strong correlation obtained between listeners’ phonetic judgments and ‘degrees’ of assimilation as defined by our proposed index of voice assimilation. This is a sufficient justification, we believe, for using that index in a quantitative investigation of voice assimilation.

The main findings of the experiments reported here can be summarized as follows. First, voice assimilation is not an all-or-none process but rather, is a graded process: there are intermediate degrees of incomplete voice assimilation. This is in line with many acoustic–phonetic studies of assimilation in other languages (place assimilation in English: Barry, 1985; Gow, 2002, 2003; Gow & Hussami, 1999; Kerswill, 1985; Nolan, 1992; Wright & Kerswill, 1989; voice assimilation in German: Kuzla, 2003). It is also in line with the previous and current research on incomplete neutralization in ‘Final Devoicing’ for languages such as Dutch and German (Dinnsen & Charles-Luce, 1984; Ernestus & Baayen, in press; Port & Crawford, 1989; Warner et al., 2004). A second, robust result is that the extent of assimilation is dependent on underlying voicing (Kohler, 1979). We consistently found that underlyingly voiceless word-final stops more easily shift toward voiced stops than voiced stops devoice. The perceptual data of Experiments 1.b and 2.b suggest that acoustic asymmetry reflects perceptual asymmetry: when extracted from their assimilatory contexts, voiceless stops are perceived as voiced more often than voiced stops are perceived as voiceless. The asymmetry we observed is also in line with perceptual data collected by other investigators. For instance, Saerens et al. (1989) examined the perception of the voicing feature in French stops. In one experiment, participants identified stop consonants from excerpts of spontaneous speech. These authors found that the percentage of correct voicing identification (i.e., corresponding to underlying voicing) was higher for voiced than for voiceless stops. This result suggests that voiced stops underwent less assimilation than voiceless stops. Another explanation for the asymmetric assimilation pattern between voiceless and voiced stops might stem from differential lexical frequencies, as we have suggested elsewhere (Snoeren & Segui, 2003). In French, there are more monosyllabic words with a voiceless than a voiced coda stop (according to “Lexique”: 2556 against 1091, word-types), suggesting that the voiceless coda is the unmarked case. When taking into account this difference, the assimilation asymmetry between voiceless and voiced stop codas could be explained in terms of “preservation” of the marked rather than the unmarked form. This is also in line with the view that speakers implicitly take lexical frequencies into account in language use (Booij, 1995; Bybee, 2001; Hay, Pierrrehumbert, & Beckman, 2003; Jurafsky, Bell, Gregory, & Raymond, 2001). To sum up, the asymmetry in assimilation between voiceless and voiced stops is probably a quite robust pattern, at least in French.
Finally, it has been suggested that phonetic neutralization can be modulated by high-level sources of information. Our data suggest that lexical confusability induced by the potential ambiguity arising from voicing confusion on word-final consonant (e.g., *soute*/soude*), or by the existence of ‘dangerous’ neighbors, can influence the extent to which word-final stops assimilate. Experiment 1 showed such a lexical ambiguity effect: potentially ambiguous words were produced with much less assimilation than unambiguous words. However, as noted previously, the experimental set was unbalanced with respect to word-final voicing and did not allow one to assess the lexical ambiguity effect independently of coda voicing. More crucially, the differences observed between potentially ambiguous and unambiguous words were probably related to the fact that speakers could notice the presence of minimal pairs in the materials, and might have tried to produce them more carefully (hence with less assimilation) than they produced the unambiguous items. Indeed, in Experiment 1, speakers practiced right away on the test sentences in order to pronounce them as fluently as possible when the recording was made. Moreover, both members of each minimal pair of potentially ambiguous words (e.g., *soute*/soude*) appeared twice in two frame sentences. In Experiment 2, subjects did not practice on the test materials. Because only assimilatory contexts were used, each member of a minimal pair appeared only once. There were only two short and semantically neutral frame sentences, and this also might have contributed to prevent speakers from detecting minimal pairs. With this more strictly controlled setting in Experiment 2, lexical ambiguity did not dramatically affect assimilation. Yet, small differences between ambiguous and unambiguous items were observed in the distribution of assimilation degrees. Likewise, voice assimilation of the word-final stop in a target word was, at least for voiceless stops, somewhat affected, by the existence of many neighbors more frequent than that target word. (The simple neighborhood density, which disregards the relative frequencies of a word and its neighbors, had no impact on assimilation degree.) So, even when speakers are not likely aware of the presence of potentially ambiguous words (Experiment 2), residual effects of lexical confusability are observed. Do they reflect a weak intention to avoid possible confusion? We would rather suggest that these effects are automatic in nature and reveal the tight interdependency of productive and perceptual processes of lexical access in speakers/listeners’ minds. In the case of word neighborhood effects, speakers are unlikely to consciously adapt their speech to avoid confusion with phonological neighbors. The specification of a more careful articulation might be intrinsically attached to the mental representations of those words that have ‘dangerous’ competitors and must be implemented in production as well as in perception.

In perception, neighborhood effects on word processing, in both the auditory and the visual word recognition domains, are very robust, provided that the neighborhood measure integrates neighbors’ frequencies (Grainger & Segui, 1990). In a study on the processing of derivation-suffixed spoken words, Meunier and Segui (1999) used a measure of competition within morphological families which is reminiscent to the NDN measure we used in Experiment 2: The morphological competition for a given word was measured as the number of the suffixed words in the same morphological family that are more frequent than that word. Lexical decision response times were longer for words with many than with few competitors. Luce (1986) used a word identification test, in which monosyllabic spoken words were presented in noise. Listeners were less accurate at identifying words with a large NDN than with a small NDN. Collectively, these studies show that lexical (and morphological) neighborhood competition hampers lexical access, which suggests that words with many competitors need additional care of articulation or acoustic
salience in order to be recognized. It is tempting to speculate that, symmetrically, they should be naturally produced more carefully. The results of Experiment 2, lend some support to that speculation.

In the present research, we addressed the issue of phonetic implementation of voice assimilation in French. We found that assimilation is often (but not always) incomplete and can be measured with an acoustic index which is perceptually meaningful. This index allowed us to show that the extent of assimilation depends on underlying voicing (voiceless stops assimilate more readily than voiced stops), whatever the speakers’ intention. Our results suggest that when speakers are presumably little concerned with avoiding potential confusion (Experiment 2), lexical confusability nonetheless somewhat affects assimilation. However, more research seems necessary to confirm that point.

The basic acoustic–phonetic data on naturally produced voice assimilation in French reported here might be helpful, we hope, in preparing perceptual experiments. Knowing more precisely how voice assimilation is phonetically implemented, we are in a better position to address the issue we raised in Section 1: Whether and how listeners recover the underlying phonemic value of assimilated segments.

Acknowledgments

This research was supported by the French Embassy, The Hague, The Netherlands with a study grant and by the French ‘Ministère de la Recherche’ with an MENRT doctoral fellowship to the first author, and with a ‘Cognitique’ (LACO 1) grant to the second and third authors. Parts of this work have been presented at the 15th ICPhS congress (August 2003) and at the 13th ESCOP conference (September 2003). We are indebted to Nicole Bacri, Willy Serniclaes, Jacqueline Vaissière, Doug Whalen, Mirjam Ernestus, and two anonymous reviewers for their comments and constructive discussion on earlier versions of this manuscript.

Appendix A. Sentences used in Experiment 1 (target words in italics)

<table>
<thead>
<tr>
<th>Non-assimilatory context</th>
<th>Assimilatory context</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Potentially ambiguous target words</td>
<td></td>
</tr>
<tr>
<td>J’ai écrit <em>bled</em> vraiment bien.</td>
<td>J’ai écrit <em>bled</em> sans erreur.</td>
</tr>
<tr>
<td>Ces <em>jades</em> ne sont pas bon marché.</td>
<td>Ces <em>jades</em> sont vraiment bon marché.</td>
</tr>
<tr>
<td>Ces <em>rides</em> n’ont rien d’extraordinaire.</td>
<td>En fait, ces <em>rides</em> sont bien naturelles.</td>
</tr>
<tr>
<td>La <em>soude</em> ne sent pas si mauvais que ça.</td>
<td>La <em>soude</em> pue vraiment fort.</td>
</tr>
<tr>
<td>J’ai écrit <em>vite</em> vraiment bien.</td>
<td>J’ai écrit <em>vite</em> sans erreur.</td>
</tr>
<tr>
<td>J’ai écrit <em>blette</em> sans erreur.</td>
<td>J’ai écrit <em>blette</em> vraiment bien.</td>
</tr>
<tr>
<td>Ces <em>jattes</em> sont vraiment bon marché.</td>
<td>Ces <em>jattes</em> ne sont pas bon marché.</td>
</tr>
<tr>
<td>Ces <em>rites</em> sont bien naturels.</td>
<td>Ces <em>rites</em> n’ont rien d’extraordinaire.</td>
</tr>
<tr>
<td>La <em>soute</em> pue vraiment fort.</td>
<td>La <em>soute</em> ne sent pas si mauvais.</td>
</tr>
<tr>
<td>J’ai écrit <em>vite</em> sans erreur.</td>
<td>J’ai écrit <em>vite</em> vraiment bien.</td>
</tr>
</tbody>
</table>
2. Unambiguous target words

On a mangé des frites croustillantes
Il a attrapé une grippe contagieuse
Il a pris des gouttes pour le nez
Les jupes courtes sont à la mode
J’ai trouvé la note plutôt salée
C’est un stock très difficile à écouler
Le comédien a un trac fou
Lisez bien le mode\textsuperscript{a} d’emploi
Il fait beau dans le sud de la France
J’ai oublié mon tube de dentifrice

Les frites de Bruxelles sont les meilleures
Il a attrapé la grippe de son frère
Il a vu une goutte de sang sur le lit
Les jupes droites lui vont bien
Chantez-moi toutes les notes de la gamme
Il a épuisé son stock de cigarettes
Il a le trac des débutants
Ce manteau est très à la mode\textsuperscript{a} cet hiver
Je n’ai jamais vu un vent du sud si fort
Ce tube passe à la radio sans arrêt

\textsuperscript{a}Although the word mode can be pronounced as the word motte after assimilation, we have decided to count mode as unambiguous because of the large difference in frequency between mode and motte (65.81 and 2.74, respectively, according to the Lexique French database).

Appendix B. Target words used in Experiment 2

<table>
<thead>
<tr>
<th>Ambiguous words</th>
<th>Voiceless</th>
<th>Voiced</th>
<th>Unambiguous words</th>
<th>Voiceless</th>
<th>Voiced</th>
</tr>
</thead>
<tbody>
<tr>
<td>bac</td>
<td>bague</td>
<td></td>
<td>botte</td>
<td></td>
<td>bide</td>
</tr>
<tr>
<td>bec</td>
<td>bègue</td>
<td></td>
<td>bouc</td>
<td></td>
<td>bob</td>
</tr>
<tr>
<td>black</td>
<td>blague</td>
<td></td>
<td>coq</td>
<td></td>
<td>bride</td>
</tr>
<tr>
<td>bock</td>
<td>bogue</td>
<td></td>
<td>coupe</td>
<td></td>
<td>cab</td>
</tr>
<tr>
<td>cote</td>
<td>code</td>
<td></td>
<td>crotte</td>
<td></td>
<td>digue</td>
</tr>
<tr>
<td>dope\textsuperscript{a}</td>
<td>daube</td>
<td></td>
<td>fac</td>
<td></td>
<td>fraude</td>
</tr>
<tr>
<td>gratte</td>
<td>grade</td>
<td></td>
<td>floc</td>
<td></td>
<td>gag</td>
</tr>
<tr>
<td>jatte</td>
<td>jade</td>
<td></td>
<td>jupe</td>
<td></td>
<td>glèbe</td>
</tr>
<tr>
<td>rate</td>
<td>rade</td>
<td></td>
<td>lac</td>
<td></td>
<td>mob</td>
</tr>
<tr>
<td>rite</td>
<td>ride</td>
<td></td>
<td>pape</td>
<td></td>
<td>rabe</td>
</tr>
<tr>
<td>soute</td>
<td>soude</td>
<td></td>
<td>soc</td>
<td></td>
<td>tube</td>
</tr>
<tr>
<td>trompe</td>
<td>trombe</td>
<td></td>
<td>steppe</td>
<td></td>
<td>vogue</td>
</tr>
</tbody>
</table>

\textsuperscript{a}The word dope was intended (and pronounced) as the English loanword, and not as the inflected form of the French verb doper. As such, dope forms a minimal pair with daube.

References


