VOICING ASSIMILATION IN JOURNALISTIC SPEECH

Hallé, P.A.[a] and Adda-Decker, M.[b]

[a] Laboratoire de Phonétique et Phonologie, [b] LIMSI
pierre.halle@univ-paris5.fr, madda@limsi.fr

ABSTRACT

We used a corpus of radio and television speech to run a quantitative study of voicing assimilation in French. The results suggest that, although voicing itself can be incomplete, voice *assimilation* is essentially categorical. The amount of voicing assimilation little depends on underlying voicing but clearly varies with segment duration and also with consonant manner of articulation. The results also suggest that voicing assimilation, though largely regressive, is not purely unidirectional.

Keywords: voicing assimilation, French, large corpus, journalistic speech

1. INTRODUCTION

We focus here on between-word assimilation such as /b/ → [p] in "…robe sale…" and do not address the issue of "within-word" assimilation, which might be considered as a case of lexicalized change rather than contextual phonological variation.

Voicing assimilation in French is often considered to be an all-or-none phenomenon and generally viewed as purely regressive. The former aspect has been questioned, however, by recent studies on French [7] as well as on other languages such as Hungarian [6], proposing that voicing assimilation is better viewed as a gradient than a categorical phonetic variation, in line with research on other assimilation processes [2, 5]. The studies on voicing used an acoustic measure of voicing degree: the proportion of voiced murmur within occlusion for stops [7], or of voiced friction within the entire consonant for fricatives [6]. As for the regressive versus progressive nature of voicing assimilation in French, we are not aware of studies that propose it is not purely regressive.

We address these two main issues using a corpus study of news speech, which somewhat differs from spontaneous speech but is closer to "natural" speech than the read speech materials used in most studies, with the notable exception of a few studies on spontaneous speech reduction [4]. At the same time, we examine the role of segment duration on assimilation: Are shorter segments more frequently assimilated? or more "strongly" assimilated (i.e., with greater assimilation degree)? Finally, throughout the study, we not only measure proportion of voiced signal as in [6-7] but examine where the voiced portions of the signal are located within segments.

2. THE JOURNALISTIC SPEECH CORPUS

The initial corpus consisted of over 100 hours of speech from radio (see [1]) and television news, together with a transcription aligned on the speech signal, containing phonemic and lexical information. About 100,000 portions were automatically extracted; each portion contained a C1#C2 sequence, where C1 is a word final consonant and C2 the initial consonant of the next word within the same sentence. We restricted consonants to the oral stops and the fricatives of French. (In French, nasals, for example, rarely induce voice assimilation processes.) Four contact conditions can be considered, according to the underlying voicing of C1 and C2 (V: voiced; NV: voiceless): NV-V and V-NV ("assimilation" conditions), NV-NV and V-V ("control" conditions). Within the C1#C2 sequences used here, there were about three times as many voiceless as voiced C1s (74,472 vs. 24,286).

Examples of extracted sequences:
excellente journée /ɛkselɑ̃**t#ʒ**urne/, *neuf décembre* /nœ**f#d**esɑ̃br/, etc.

These sequences contained about one second of speech around the C1#C2 contact, together with the segmental labels and time locations produced by the LIMSI automatic alignment system.

3. MEASUREMENTS

We followed [7] and [6] for the measurement of a voicing ratio for stop and fricative consonants: the proportion of voiced signal within stop occlusion or fricative constriction. The labeling information for the C1#C2 sequences provides the boundary locations of both C1 and C2. We used the F0 extraction module in Praat [3], with a 3 ms time-step and otherwise standard settings, to determine voicedness, thus, for both consonants, a *voicing ratio* (henceforth, v–ratio) in the 0-1 range. The phonetic-acoustic alignment system used a 10 ms analysis step, and produced segment durations not shorter than 30 ms. Because limited accuracy in segment boundary location is expected for the shortest segments, we examined v-ratios separately for four C1#C2 sequence duration ranges: 60-120, 120-180, 180-240, and more than 240 ms. For the shortest durations (60-120 ms), accuracy might not be optimal and the results must be considered with caution. For very long durations, accuracy might also be lacking in that C1#C2 may encompass a pause or a schwa-like vowel, integrated by the alignment system into either C1 or C2. The most reliable results are thus expected for the intermediate ranges, from 120 to 240 ms. In spite of these potential shortcomings, which are inherent to any automatic alignment system, the analyses proved to be rather consistent and homogeneous.

In the case of voiceless C1s (especially stops) preceded by a vowel, the closure portion overlaps with the voicing lag of this vowel, that is, it usually begins with a short voiced portion decreasing in amplitude which reflects the vowel offset rather than a voiced closure portion. We therefore always compare v–ratios between an assimilatory situation and a corresponding non-assimilatory, "control" situation to assess voicing assimilation.

Voicing ratio tells us whether a consonant is fully voiced, fully voiceless, or incompletely voiced. In the latter case, the question arises as to which portion of the consonant is voiced. We distinguish four configurations according to whether the voiced portion is located at the left edge of the consonant, at the right edge, scattered at both edges, or lies in the middle of the consonant. Naive intuition about regressive assimilation suggests the *right edge* of C1 should be affected by a following C2 with a different underlying voicing. Yet, stops and fricatives might behave differently in this respect. In voiceless stops, the occlusion usually begins with the voicing lag of a preceding vowel. It might be the case that this voiced portion extends to the right in a [+voice] assimilation context. In case it does not extend to the entire occlusion, we may conclude to partial assimilation.

All the measurements were run automatically on the set of annotated C1#C2 sequences, producing, for each C1 or C2, total duration, voiced duration, beginning and end locations of the voiced portion. Based on these data, the sequences were categorized as fully voiced, fully voiceless, or partially voiced (with one of four configurations).

4. RESULTS

4.1. C1: voicing configurations and v-ratios

The right-edge, both-edge, or mid-part voicing configurations were quite marginal in frequency (less than 2% in average), for stops and fricatives pooled. We thus focus, in the following, on the left-edge partial voicing configuration and the fully voiced or fully voiceless cases.

Tables 1 and 2 show the overall distribution of C1s according to voicing configuration in the four C1#C2 contact situations.

Table 1: distribution and v-ratio of C1 according to C1 voicing configuration: all (fully voiced), left-edge voiced, and none (fully voiceless): NV-V vs. NV-NV.

		voicing configuration		
C1#C2		all	left	none
NV-NV	%	9	64	20
	v-ratio	*1*	*.36*	*0*
NV-V	%	58	23	9
	v-ratio	*1*	*.35*	*0*

Table 2: same as Table 1 for V-NV vs. V-V.

		voicing configuration		
C1#C2		all	left	none
V-V	%	83	11	2
	v-ratio	*1*	*.56*	*0*
V-NV	%	35	52	10
	v-ratio	*1*	*.50*	*0*

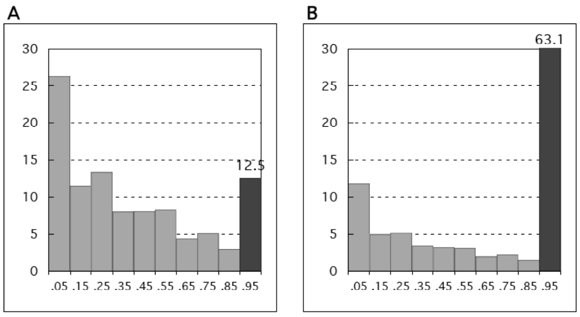
In the NV-V situation, the frequency of 'fully voiced' cases increases by 49% compared to the NV-NV situation, compensating for the decrease in frequency of the 'left-edge voicing' cases (-41%) and the 'fully voiceless' cases (-11%).

In the V-NV situation, the frequency of fully voiced cases decreases by 48% compared to the V–V situation, compensating for the increase in frequency of the 'left-edge voicing' cases (+41%) and the 'fully voiceless' cases (+8%).

The two directions of assimilation thus seem fairly symmetrical for these raw count variations.

One issue of interest is of whether the data speak for categorical rather than partial voice assimilation. The former view entails an all-or-none change: in the NV-V compared to the NV-V condition, for instance, the *number* of fully voiced C1s should increase (full change cases), and the non-fully voiced C1s correspond to the no-change cases. The distribution of the non-fully voiced C1s thus should not vary substantially (from NV-NV to NV-V) if voice assimilation is categorical: some of these C1s would not change, some others would switch to fully voiced. Partial voice assimilation, on the other hand, entails a shift toward greater v–ratio for *all* C1s. This would mean a shift in the v–ratio distribution for non-fully voiced C1s. As can be seen in Fig. 1, this is clearly not the case.

Figure 1: Distributions of C1 v-ratios for (A) NV-NV and (B) NV-V conditions. The black bar corresponds, roughly, to fully voiced C1s (partly shown for B).

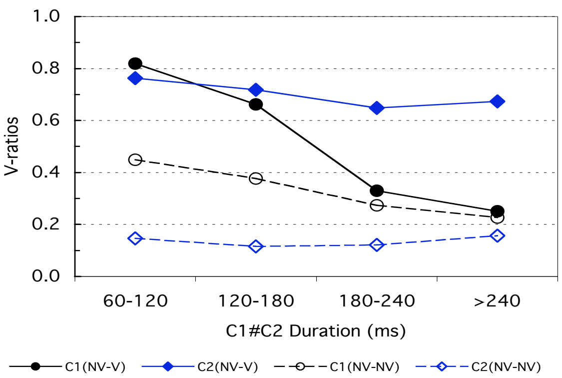


The distributions of non-fully voiced C1s are virtually identical in the NV-NV and NV-V conditions, $\chi^2(8) = 1.22$, $p = .996$, n.s., suggesting full switch to fully voiced in some cases and no switch at all in the remaining cases. The comparison between the V-NV and V-V conditions essentially yields a symmetrical pattern: a transfer from the fully voiced to the non-fully voiced (including fully voiceless) categories, rather than a shift in v–ratio distribution, hence also suggests categorical rather than partial voice assimilation. (There are, however, differences between fricatives and stops for the [-voice] assimilation: fricatives seem to be devoiced less categorically than stops.)

4.2. C1 and C2 v-ratios according to duration

As we noted, the data may be less reliable for extreme C1#C2 durations. In the following, we examine how v-ratios of C1 and C2 vary according to sequence duration and contact condition. The results are shown in Figs 2-3.

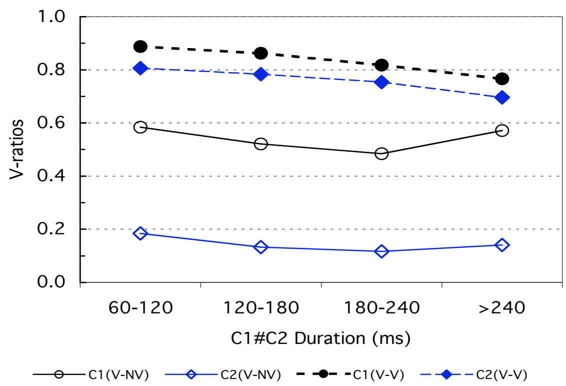
Figure 2: v-ratios for C1 (circles) and C2 (diamonds); NV-V/NV-NV: solid/dashed lines, filled/empty marks.



In the NV-V and NV-NV conditions, v-ratios for C2 (diamonds) vary little with sequence duration. Those for C1 (circles) vary much more, especially for voiceless C1 followed by voiced C2. In the NV-NV control condition, v–ratios for C1 decrease with sequence duration, yet remain above those for C2. Because these v–ratios are due to the voicing lag of a preceding vowel in most cases, the observed variation suggests that, for longer C1s, the vowel "voiced tail" simply is proportionally shorter. The smaller v-ratios for C2 are presumably due to the absence of a preceding vowel (and its voiced tail) in most cases.

In the NV-V assimilation condition, v–ratios for C1 decrease dramatically with sequence duration. For the short duration ranges (60-180 ms), C1 and C2 have equivalent v-ratios. That is, C1s are voice-assimilated. Little or no assimilation occurs for long C1#C2 sequences, in which a vocalic release, or a pause, may occasionally occur after C1. As we discussed, a vocalic release could be included in either C1 or C2 by the alignment system. In this case, however, the inaccurate labeling would overestimate rather than underestimate v–ratios. The low v–ratios for C1 in the NV-V condition for long C1#C2 sequences are thus more likely due to an intervening voiceless pause.

Figure 2: v-ratios for C1 (circles) and C2 (diamonds); V-NV/V-V: solid/dashed lines, empty/filled marks.



In the V-NV and V-V conditions, there is much less variation overall than in the other conditions. In the V-NV condition, C1 v-ratio is slightly larger for the longest C1#C2 duration, reflecting, perhaps, a lesser degree of devoicing assimilation. Yet, as discussed above, it also might be due to inaccurate labeling when an untranscribed epenthetic vowel is produced between C1 and C2.

For the short duration ranges (60-180 ms), the results show that assimilation mainly affects C1: the v–ratios of voiceless C1s increase by about .33

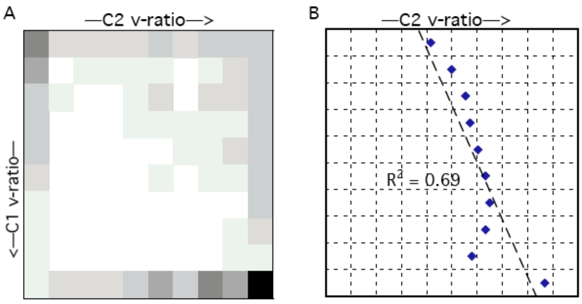
and those of voiced C1s decrease by about .34 in the assimilation condition compared to the control condition. In the following, we focus on the 120-180 ms duration range because it should be less prone to labeling errors than the other ranges (due to pause or schwa insertions for long durations), and because it is the most representative (45.5% of all sequences, against 38.8, 12.0, and 3.7% for the 60-120, 180-240, and >240 ranges, respectively).

4.3. V-ratio distributions (120-180 ms range)

We examined thus far only global aspects of voicing assimilation, yet it seems that most of the voicing changes in C1#C2 sequences concern C1: this assimilation is mainly regressive in French.

Is there some progressive component? To address this question, we used the distributions of the C1#C2 sequences according to C1 and C2 v–ratios, in the four contact conditions, as illustrated in Fig. 4 for the NV-V condition.

Figure 4: NV-V: (A) distribution of C1#C2 according to C1 x C2 v-ratios (frequencies represented by gray levels); (B) correlation between C1 and C2 v-ratios.



In this condition, about half of the sequences fall in the black cell for C1 and C2 v-ratios above .9. But there are 5% sequences that fall in the cell of the opposite corner (v-ratios below .1): these sequences may correspond to progressive (devoicing) assimilation. We may ask whether there is some covariation between C1 and C2 v-ratios The mean C2 v-ratio per C1 v-ratio interval (computed over 10 intervals) correlates significantly with the center v-ratio of that interval: that is, the more C1 is voiced, the more C2 is voiced. Such correlation is found in all contact conditions, at least at the $p<.05$ level ($r(8)>0.65$), except in the V-V control condition. But for V-V, there is little variation in v–ratio for both C1 and C2 (Fig. 3). The influence of C1 on C2 can also be seen in comparing the data in Figs. 2-3: voiced C2s have lower v-ratios after voiceless than voiced C1s although the differential (~.05) is much smaller than for voiced C1s in NV–V compared to NV-NV (~.33); voiceless C2s, however, are hardly influenced by C1s.

Stops and fricatives somewhat differ. There is less overlap with a preceding vowel offset in fricatives, thus more leeway for assimilation: variation in v-ratio is greater for fricatives (~.40) than for stops (~.30) by about .10.

5. SUMMARY AND CONCLUSIONS

In this study, we report some essential aspects of voicing assimilation in French, based on a large corpus of journalistic speech. Using "voicing ratio" as a measure of voicing degree, incomplete voicing of C1 was found in all situations of C1#C2 consonant contact, and within the category of partially voiced C1s, consonants differing by their underlying voicing differed by their voicing ratio. Hence, voicing itself is fairly gradient. But this does not entail that voicing *assimilation* also is gradient, as we suggest in section 4.1. Rather, voicing assimilation might be categorical in French, contrary to claims made in [7]. For the two directions of assimilation, the data also show equivalent amounts of change in voicing ratio, and symmetrical transfers between voicing configuration categories. We finally find that voicing assimilation is not purely unidirectional: at least at the acoustic level, some interaction seems to take place between C1 and C2 "degrees" of voicing. This point, however, might need further clarification, and should be examined, in particular, across different C1#C2 durations. The same applies to the differences between stops and fricatives that we only briefly mentioned.

6. REFERENCES

[1] Adda-Decker, M., Hallé, P. 2007. Bayesian framework for voicing alternation and assimilation studies on large corpora in French, *Proc. 16th ICPhS*, Saarbrücken.
[2] Barry, M. C. 1985. A palatographic study of connected speech processes. Cambridge *Papers in Phonetics & Experimental Linguistics* 4, 1-16.
[3] Boersma, P., Weenink, D. 1992-2004. Praat: A system for doing phonetics by computer [Computer program]. http://www.praat.org.
[4] Duez, D. 1995. On spontaneous French speech: aspects of the reduction and contextual assimilation of voiced stops. *Journal of Phonetics* 23, 407-427.
[5] Gow, D. W., Hussami, P. 1999. Acoustic modification in English place assimilation. *JASA* 106, 2243 [abstract].
[6] Gow, D. W., Im, A. M. 2004. A cross-linguistic examination of assimilation context effects. *Journal of Memory and Language* 51, 279-296.
[7] Snoeren, N., Hallé, P., Segui, J. 2006. A voice for the voiceless: Production and perception of assimilated stops in French. *Journal of Phonetics* 34, 241-268.