*INTERSPEECH 2019*
September 15–19, 2019, Graz, Austria

# "*Gra*[f]*e*!" Word-final devoicing of obstruents in Standard French: An acoustic study based on large corpora

*Adèle Jatteau*[1], *Ioana Vasilescu*[1], *Lori Lamel*[1], *Martine Adda-Decker*[1,2], *Nicolas Audibert*[2]

[1]LIMSI-CNRS, Université Paris-Saclay, Orsay, France
[2]LPP, CNRS-Université Sorbonne Nouvelle, Paris, France
{jatteau, vasilescu, lamel, madda} @limsi.fr, nicolas.audibert@sorbonne.nouvelle.fr

## Abstract

This study investigates the tendency towards word-final devoicing of voiced obstruents in Standard French, and how devoicing is influenced by domain, speech style, manner and place of articulation. Three large corpora with automatic segmentations produced by forced alignment are used: ESTER, ETAPE and NCCFr. A voicing-ratio is established for each obstruent via F0 extraction in Praat, and the percentage of fully voiced segments is computed. We find a salient pattern of devoicing before pause, with no clear effect of speech style. Fricatives devoice more than stops, and posterior fricatives devoice more than anterior ones. Since voicing plays a central role in the cross-linguistic pattern of word-final [voice] neutralisation, this study gives insight into the potential phonetic precursors of this process.
**Index Terms**: voicing, large corpora, forced alignment, Standard French, acoustics

## 1. Introduction

Final devoicing (FD) is the process whereby (contrastively) voiced consonants are devoiced in domain-final position, as in Russian *Youtu*[p]. Many factors converge to suggest that FD is a "natural" process: it is widely attested cross-linguistically [1], constitutes a frequent sound change [2, pp. 184-186], and appears regularly in L1 and L2 acquisition [3]. Several sources for FD have been proposed in the phonetic literature: lack of C-V transition and its cues to the voicing contrast [4]; anticipation of the glottal opening for breathing [5]; utterance-final decrease of subglottal pressure yielding voicing offset prior to the obstruent release [6]; failure of voicing production, as well as perception, in utterance-final lengthening [1, 7]. These phonetic sources predict that a variable FD effect should be found in languages which do not present a phonologised process of final neutralisation. To test this hypothesis, this paper studies the voicing alternations of obstruents in word-final position in three large corpora of Standard French.

Standard French is a good candidate because it robustly allows obstruents to contrast for voicing in word-final position: e.g. *bac* 'ferry' / *bague* 'ring'. This is not the case of all the variants of the language: FD is reported as a regular process in Belgian French [8], and is a well-known feature of the French adjacent region, in the North [9, 10], as well as in Alsace [11, 10], in contact with German. However, these are regionally marked variants, and FD is not usually observed for Standard French. Beyond these regional variants, nevertheless, a few studies also report a minor tendency towards FD in the standard variety [12, 13]. Jatteau et al. [14] investigate FD in two large corpora of Standard French, focusing on fricatives, and report a devoicing effect in pre-pausal position. The present paper expands this investigation to all obstruents and to a third corpus, in order to get a finer view of the effect of speech style. It also uses the proportion of the consonant which is voiced (via F0 detection) instead of phone categories assigned by the forced alignment.

Examining this question in large corpora allows us to quantify the variable tendency towards FD under less supervised settings than laboratory recordings. The size of the corpora also enables us to describe it more precisely. In addition to the main research question above, four hypotheses are investigated. First, since the phonetic sources listed above refer to the utterance-final position, we should expect to find a tendency towards FD before pause, rather than in all word-final contexts (H1) [1, 5]. Second, variation is expected primarily in less formal, spontaneous styles of speech. We should therefore find more devoicing in this register (H2). Third, the process should be sensitive to the manner of articulation of the obstruents: since the high degree of oral pressure required by fricatives conflicts with the low degree of oral pressure required for voicing [15], voiced fricatives should be more prone to devoicing than voiced stops (H3). Finally, posterior obstruents should present more devoicing than the anterior ones (H4): with a smaller vocal tract, it is more difficult to maintain the pressure differential across the glottis [7].

## 2. Data and methodology

### 2.1. Corpora and alignment

Three manually transcribed corpora of Standard French are investigated: ESTER, which consists of 90 hours of radio and TV broadcast news (BN) [16]; ETAPE, which contains 42 hours of radio and TV BN and entertainment [17]; and NCCFr, which comprises 36 hours of conversation between friends [18]. These three corpora are associated with three different styles of speech: formal, semi-formal and informal.

The data were segmented with forced alignment using the LIMSI's automatic speech recognition system [19]. Forced alignment takes as input the audio files and their transcription, and returns the word and phone boundaries as well as their labels. We extracted from the aligned transcription the list of word-final voiced obstruents, for a total of about 195000 tokens. These include schwa-final words (Cə#) along with obstruents in absolute final position (C#).

To ensure the data are representative of the phenomenon of interest, we listened to a sample of 500 extracts distributed over the whole dataset, and applied a series of filters. First, in order to restrict the study to Standard French, we filtered out the audio files from *Radio Télévision Maroc* and *Radio France International* (in the ESTER corpus), since they contain many portions of African and Maghrebine French. This operation suppressed about 13500 tokens. We also eliminated from all corpora words consisting of one obstruent (such as the preposition *d'* for *de*), which are not phonologically word-final and are extremely fre-



http://dx.doi.org/10.21437/Interspeech.2019-2329

quent (136000 tokens). Further cleaning excluded incomplete words, interjections, allomorphs (e.g. *plus* /ply/ ~ /plys/), and loanwords which allow different pronunciations (e.g. *Gomez* [z]/[s]; 12500 tokens).

Second, in order to eliminate grossly misaligned cases, segments longer than the mean duration + twice the standard deviation were filtered out. This threshold was calculated for each phoneme within each corpus (for instance /b/ in ESTER), in order to respect the relative durations of the different phonemes and speech styles. This filtered out 1350 tokens.

These operations yielded a database of 30872 word-final voiced obstruents (/b, d, g, v, z, ʒ/): 25986 tokens in absolute final position (C#) and 4986 followed by a schwa (Cə#). These obstruents were sorted in 5 categories depending on the following context: the next word begins with

- a voiceless obstruent 'NVObst' ex. *arrive tôt*
- a voiced obstruent 'VObst' ex. *arrive demain*
- a sonorant 'Son' ex. *arrive lundi*
- a vowel 'Vow' ex. *arrive avant*
- or the obstruent is followed by a pause (silence or breath, ex. *arrive* ##).

As Fig. 1 shows, the data are not evenly distributed across these contexts: word-final voiced obstruents are most often followed by obstruent or vowel, and least often by sonorant or pause.

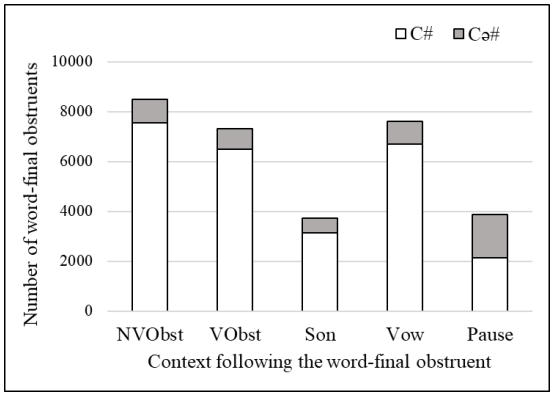


Figure 1: *Number of word-final obstruents as a function of the following context.*

### 2.2. Extraction of the voicing-ratio

The proportion of voiced signal in these segments was measured using the F0 extraction module in Praat [20]. This method follows [21] and [22] and allows us to extract a voicing ratio (henceforth *v-ratio*) in the 0%-100% range. V-ratio is defined as the number of points detected as voiced by Praat, divided by the total number of measurement points (20).

Examination of this v-ratio shows that it has a skewed distribution: in many contexts, more than 50% of the obstruents are fully voiced. Fig. 2 illustrates the distribution of the v-ratio values of word-final obstruents in absolute final position before pause. In this position, 31% of the word-final obstruents are fully voiced. Because of this asymmetry, we investigate the voicing alternations of word-final obstruents in two different ways. First, we focus on the percentage of fully voiced obstruents (v-ratio = 100%), as opposed to the partially voiced ones (v-ratio < 100%). Second, the range of v-ratio values and their distribution are used for some illustrative examples, to highlight the degree of variation which appears in the results.

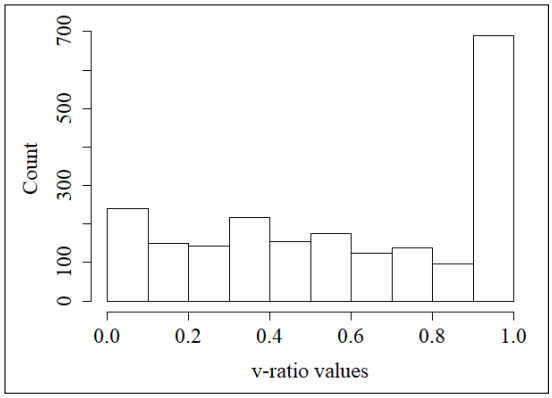


Figure 2: *Distribution of the v-ratios of word-final obstruents in absolute final position before pause (number of tokens: 2129).*

## 3. Results

### 3.1. Evidence of utterance-final FD

We calculated the percentage of fully voiced word-final obstruents in each of the 5 contexts for the three corpora pooled together. Table 1 shows that when all word-final obstruents are taken into account, two contexts trigger a decrease in the proportion of fully voiced obstruents. The first and strongest one is before a voiceless obstruent: only 29% of the obstruents in this context are fully voiced. This is due to the laryngeal assimilation effect (as in *arri*[f]*e tôt*) [21, 22]. The second is before pause: only 52% of the obstruents before pause are fully voiced. This is evidence for a FD pattern in utterance-final position, which confirms our H1. Before a voiced obstruent, a sonorant or a vowel, on the other hand, the percentage of fully voiced segments is high (respectively 80, 75 and 74%).

When obstruents followed by a schwa are considered separately, the percentage of fully voiced obstruents in the Cə# condition remains high, with little variation. Unsurprisingly, obstruents followed by a schwa do not contribute to the final devoicing pattern. When only obstruents in absolute final position are examined, the proportion of fully voiced segments before pause decreases to 31%. In order to focus on the devoicing pattern, schwa-final words are excluded from the analyses below.

Table 1: *Percentage of fully voiced word-final obstruents as a function of the following context and the presence of a schwa.*

| | NVObst | VObst | Son | Vow | Pause |
|---|---|---|---|---|---|
| C# + Cə# | 29 | 80 | 75 | 74 | 52 |
| Cə# only | 73 | 74 | 73 | 69 | 77 |
| C# only | 24 | 81 | 75 | 74 | 31 |

Fig. 3 and 4 provide two different visual illustrations of the voicing alternations in our corpora for the 5 contexts: Fig. 3 shows the percentage of fully voiced segments, and Fig. 4 illustrates the distribution of the v-ratio itself. Fig. 4 shows that there is a lot of variation in the two devoicing contexts, with an interquartile range of 70% before voiceless obstruent and 70% before pause. Anecdoctically, note that there are more fully voiced obstruents before voiced obstruent (81% in Fig. 3) than before sonorant or vowel, suggesting a phonetic voicing assimilation effect even for phonologically voiced consonants. This

point stands however beyond the scope of this paper. Since the FD pattern is found in pre-pausal position, the remainder of this paper focuses on this context.

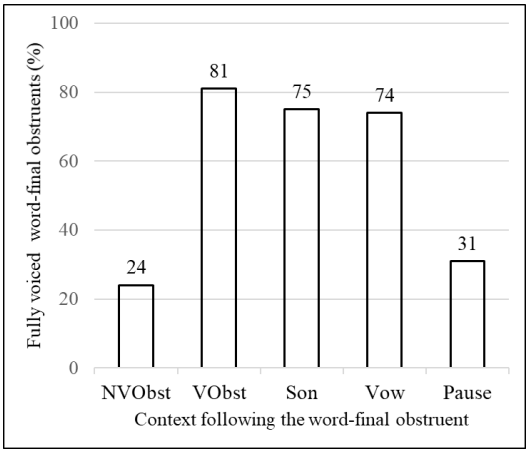


Figure 3: *Percentage of fully voiced word-final obstruents in absolute final position as a function of the following context.*

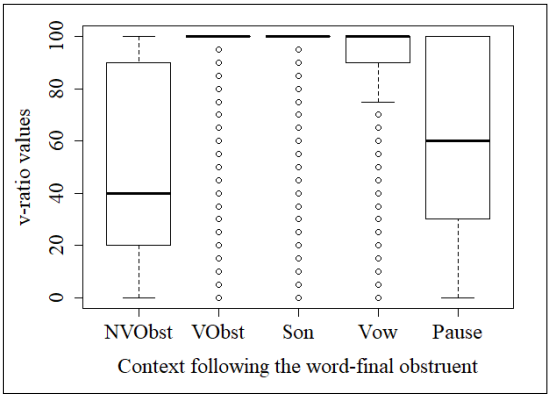


Figure 4: *V-ratio of word-final obstruents in absolute final position as a function of the following context.*

### 3.2. Effect of speech style

When the three corpora and their associated speech styles are considered separately, we find that the semi-formal speech style of ETAPE stands out with the highest rate of fully voiced obstruents before pause (38%), whereas the formal speech style of ESTER and the casual speech style of NCCFr show comparable rates (respectively 28% and 30%; cf. Table 2). The difference between ETAPE and the other corpora is significant (ETAPE/ESTER : $\chi^2(1) = 17.02$, $p < 0.001$; ETAPE/NCCFr $\chi^2(1) = 8.58$, $p = 0.003$), but ESTER and NCCFr have similar rates of fully voiced tokens before pause ($\chi^2(1) = 0.7$, $p = 0.4$).

These results are not in line with our H2 in two ways: NCCFr was expected to show more devoicing than ESTER, and ETAPE was expected to be intermediate between ESTER and NCCFr. Regarding the second point, the higher proportion of fully voiced segments in ETAPE might be explained by the more heterogeneous recording conditions of the corpus; portions of dialog overlaps in debates and broadcast interviews may artificially raise the frequency of F0 detection. On the other hand, the equivalence between the most formal and the most informal styles of speech is an interesting result. It suggests that the reduction of voicing in pre-pausal position may not be sociolinguistically coloured, and might reflect physiological effects.

### 3.3. Effect of manner of articulation

The comparison of stops and fricatives shows that before pause, there are much fewer fully voiced fricatives than fully voiced stops. Table 2 shows that this holds for all corpora pooled together as well as for each of them separately ($p < 0.001$ in all cases). This tendency is not specific to the pre-pausal context: when all contexts are considered together, 50% of the fricatives are fully voiced, vs. 73% of the stops (in absolute final position). This results confirms our H3: in word-final position, fricatives are more prone to devoicing than stops.

Table 2: *Percentage of fully voiced word-final obstruents in absolute final position before pause as a function of manner of articulation and corpus.*

| | **Stops** | **Fricatives** | **All obstruents** |
|---|---|---|---|
| All corpora | 43 | 27 | 31 |
| ESTER | 40 | 24 | 28 |
| ETAPE | 50 | 33 | 38 |
| NCCFr | 40 | 24 | 30 |

### 3.4. Effect of place of articulation

Finally, the last research question we address is the effect of place of articulation. Looking first at the stops in Fig. 5, we can see that the rate of fully voiced stops before pause is not affected by their place of articulation: contrary to H4, it is not the case that velars devoice more than alveolars or labials ($\chi^2(2) = 0.32$, $p = 0.84$). For fricatives on the other hand, the expected effect is clear: the more posterior the fricative, the less fully voiced segments we find. More precisely, /ʒ/ devoices more than /z/ ($\chi^2(1) = 35.01$, $p < 0.001$), and /z/ devoices more than /v/ ($\chi^2(1) = 18.79$, $p < 0.001$).

Fig. 6 shows that the same tendencies are reflected when the whole range of v-ratio values is taken into account: stops show comparable distributions (none of the comparisons is significant with Mann-Whitney-Wilcoxon tests), but fricatives show a significant decrease from anterior ones (less devoiced) to posterior ones (more devoiced).

Finally, note that /ʒ/ is also the obstruent with the lowest rate of fully voiced tokens in general, regardless of the context (Table 3). The pre-pausal position therefore amplifies a general tendency for /ʒ/ to be partially devoiced word-finally.

Table 3: *Percentage of fully voiced segments in absolute final position as a function of phoneme and corpus (all contexts).*

| | /b/ | /d/ | /g/ | /v/ | /z/ | /ʒ/ |
|---|---|---|---|---|---|---|
| All corpora | 74 | 74 | 69 | 58 | 52 | 39 |
| ESTER | 78 | 75 | 69 | 67 | 52 | 39 |
| ETAPE | 76 | 80 | 74 | 63 | 57 | 46 |
| NCCFr | 64 | 63 | 64 | 47 | 42 | 30 |

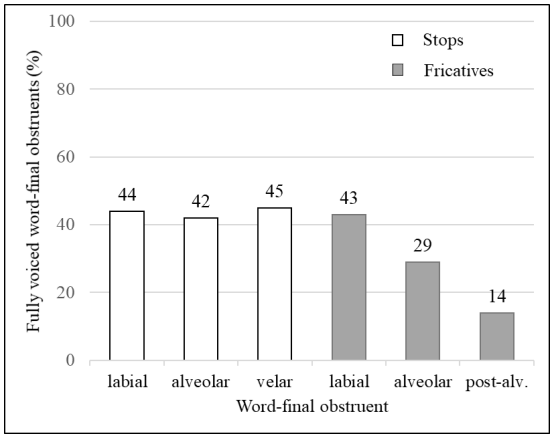


Figure 5: *Percentage of fully voiced word-final obstruents in absolute final position before pause as a function of the place of articulation.*

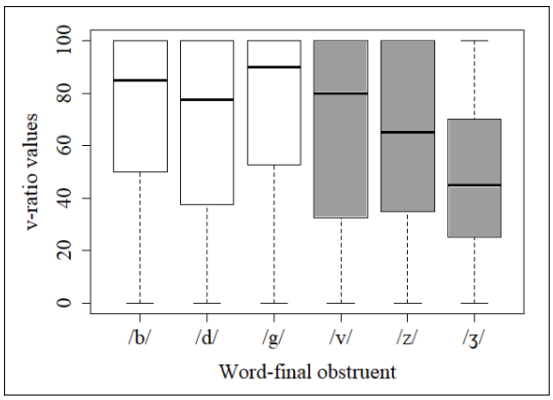


Figure 6: *V-ratio of word-final obstruents in absolute final position before pause as a function of the phoneme.*

## 4. Conclusions

The results of this study confirm that, once the effect of laryngeal assimilation is set apart, there is a variable tendency for devoicing voiced obstruents in word-final position in Standard French. In confirmation of our H1, this effect is limited to the pre-pausal context: before a silence or a breath, only 31% of the word-final obstruents in absolute final position are fully voiced. This variable devoicing at the end of utterances is consistent with the classical diachronic scenario for edge processes, which begin in utterance-final position before generalising to all word-final positions [23].

Our H2 predicted that speech style affects variation, in the sense that more variation should be found in casual, informal types of speech. Our results do not confirm this hypothesis: the most informal style of speech (conversation in NCCFr) does not devoice more than the most formal style of speech (broadcast news in ESTER). This suggests that the pre-pausal reduction of voicing does not depend (only) on sociolinguistic parameters, at least not on the overall speech style we globally associated with each corpus, and that it might result from physiological constraints. The fact that ETAPE shows less devoicing than the other two corpora, although we expected an intermediate pattern, might be linked to the more heterogeneous recording settings of the corpus, with overlapping speech artificially raising the F0 detection. This question requires further analysis of the results and may call for additional filtering.

H3 predicted that fricatives devoice more than stops, because they impose contradictory aerodynamic requirements to the oral and infra-oral pressures. This hypothesis is confirmed in our results: fricatives show a lower rate of fully voiced tokens than stops (27% vs. 43%). Finally, posterior fricatives devoice more than the anterior ones, as predicted by H4: /ʒ/ devoices more than /z/, which devoices more than /v/. This is a special case of a more general tendency for /ʒ/ in particular to be only partially voiced. Stops however do not show any effect of the place of articulation.

These results can be compared to our previous results on this research project. In [14], we studied the voicing alternations of word-final fricatives (in absolute final position only) using a different methodology: voicing was evaluated as a binary decision, taken by the automatic aligner itself based on its acoustic models for voiced and voiceless fricatives. For instance, the aligner was allowed, for the word *grève*, to transcribe it [gʁɛv] or [gʁɛf], depending of whether the fricative more closely resembled its acoustic models for /v/ or for /f/. The present study and the previous one converge to the main result: there is prepausal devoicing in Standard French. However, [14] found a different effect of speech style: NCCFr showed more devoicing than ESTER. [14] also found a different effect of place of articulation: /v/ devoiced more than /z/ and /ʒ/. These differences may at least partially be due to the fact that the acoustic models of the aligner use much more information to diagnose the [voice] contrast than the sinusoidal shape of the waveform. So do humans: a vast phonetic literature has shown that this contrast relies on an array of different cues, such as preceding vowel duration and consonant duration, and not only on the presence or not of glottal vibration. A complete study of FD should therefore investigate whether these other cues are also variably weakened in languages such as French.

Nevertheless, glottal vibration and its correlate, VOT, have been shown to be the main cue for voicing in French [24] and in other languages [25]. Moreover, many studies have shown that in languages which present a phonologised FD effect, such as Catalan or Russian, the neutralisation of the [voice] contrast is acoustically incomplete, and glottal vibration is the most affected cue [26, 27]. On the one hand, this means that glottal vibration is an important factor in the process of FD, so that the present study on the variation of voicing *per se* constitutes an interesting testimony of a potential phonetic precursor to the phonologisation of FD. On the other hand, this central role of voicing in FD is consistent with the fact that three of the five phonetic sources for FD listed in §1 (anticipation of the breathing configuration, decrease in subglottal pressure utterance-finally, lengthening of the consonant duration) actually affect glottal vibration, rather than the other cues to the voicing contrast. An open question is then how the other cues to the voicing contrast get reduced in phonologised FD.

## 5. Acknowledgements

This work was partially supported by the *Maison des Sciences de l'Homme* Paris-Saclay through a Maturation grant, and by the French National Agency for Research as part of the SALSA project under grant ANR-14-CE28-0021.

# 6. References


[1] J. Blevins, "Theoretical synopsis of Evolutionary Phonology," *Theoretical Linguistics*, vol. 32, no. 2, pp. 117–166, 2006.

[2] M. Kümmel, *Konsonantenwandel: Bausteine zu einer Typologie des Lautwandels und ihre Konsequenzen für die vergleichende Rekonstruktion*. Wiesbaden: Reichert Verlag, 2007.

[3] E. Broselow, "Laryngeal contrasts in second language phonology," in *Phonological Typology*, L. M. Hyman and F. Plank, Eds. Berlin ; Boston: De Gruyter Mouton, 2018, pp. 312–340.

[4] D. Steriade, "Alternatives to the syllabic interpretation of consonantal phonotactics," in *Proceedings of the 1998 Linguistics and Phonetics Conference*, O. Fujimura, B. D. Joseph, and B. Palek, Eds. The Karolinum Press, 1999, pp. 205–242.

[5] S. Myers, "Final devoicing: Production and perception studies," in *Prosody matters: Essays in honor of Elisabeth Selkirk*, T. Borowsky, S. Kawahara, T. Shinya, and M. Sugahara, Eds. Equinox, 2012, pp. 148–180.

[6] J. Westbury and P. Keating, "On the naturalness of stop consonant voicing," *Journal of Linguistics*, vol. 22, pp. 145–166, 1986.

[7] J. J. Ohala, "Aerodynamics of phonology," in *Proceedings of the Seoul International Conference on Linguistics*. Seoul: Linguistic Society of Korea, 1997, pp. 92–97.

[8] P. Hambye, "La prononciation du français contemporain en Belgique : variation, normes et identités," Ph.D. dissertation, Université Catholique de Louvain, 2005.

[9] T. Pooley, "Word-final consonant devoicing in a variety of working-class French - a case of language contact?" *Journal of French Language Studies*, vol. 4, no. 2, pp. 215–233, 1994.

[10] R. Temple, "Old wine into new wineskins. A variationist investigation into patterns of devoicing in plosives in the Atlas linguistique de la France," *Transactions of the Philological Society*, vol. 98, pp. 353–294, 2000.

[11] J.-P. Montreuil, "Multiple opacity in Eastern Regional French," in *Romance Linguistics 2009: Selected Papers fron the 39th Linguistic Symposium on Romance Languages (LSRL), Tucson, Arizona*, S. Colina, A. Olarrea, and A. Carvalho, Eds. Amsterdam: John Benjamins, 2010, pp. 153–166.

[12] H. Walter, *La phonologie du français*. Paris: Presses Universitaires de France, 1977.

[13] R. Temple, "Phonetic and sociophonetic conditioning of voicing patterns in the stop consonants of French," *14th International Congress of Phonetic Sciences*, pp. 1409–1412, 1999.

[14] A. Jatteau, I. Vasilescu, L. Lamel, and M. Adda-Decker, "Final devoicing of fricatives in French: Studying variation in large-scale corpora with automatic alignment," in *19th International Congress of Phonetic Sciences*, 2019, pp. 295–299.

[15] J. J. Ohala, "The origin of sound patterns in vocal tract constraints," in *The production of speech*, P. MacNeilage, Ed. New York: Springer-Verlag, 1983, pp. 189–216.

[16] S. Galliano, E. Geoffrois, D. Mostefa, K. Choukri, J.-F. Bonastre, and G. Gravier, "The ESTER phase II evaluation campaign for the rich transcription of French broadcast news," in *9th European Conference on Speech Communication and Technology*, 2005.

[17] G. Gravier, G. Adda, N. Paulson, M. Carré, A. Giraudel, and O. Galibert, "The ETAPE corpus for the evaluation of speech-based TV content processing in the French language," in *LREC – 8th International Conference on Language Resources and Evaluation*, 2012.

[18] F. Torreira, M. Adda-Decker, and M. Ernestus, "The Nijmegen corpus of casual French," *Speech Communication*, vol. 52, no. 3, pp. 201–212, 2010.

[19] J.-L. Gauvain, L. Lamel, and G. Adda, "The LIMSI broadcast news transcription system," *Speech communication*, vol. 37, no. 1-2, pp. 89–108, 2002.

[20] P. Boersma and D. Weenink, "Praat: Doing phonetics by computer," 2017. [Online]. Available: <http://www.praat.org/>

[21] N. Snoeren, P. Hallé, and J. Segui, "A voice for the voiceless: Production and perception of assimilated stops in French," *Journal of Phonetics*, vol. 34, pp. 241–268, 2006.

[22] P. Hallé and M. Adda-Decker, "Voicing assimilation in journalistic speech," in *16th International Congress of Phonetic Sciences*, 2007, pp. 493–496.

[23] R. Bermúdez-Otero, "Amphichronic explanations and the life cycle of phonological processes," in *The Oxford Handbook of Historical Phonology*, J. C. Salmons and P. Honeybone, Eds. Oxford: Oxford University Press, 2015, pp. 374–399.

[24] W. Serniclaes, "Etude expérimentale de la perception du trait de voisement des occlusives du français," Ph.D. dissertation, Université Libre de Bruxelles, 1987.

[25] L. Lisker and A. S. Abramson, "A cross-language study of voicing in initial stops: Acoustical measurements," *Word*, vol. 20, pp. 384–422, 1964.

[26] O. Dmitrieva, A. Jongman, and J. Sereno, "Phonological neutralization by native and non-native speakers: The case of Russian final devoicing," *Journal of Phonetics*, vol. 38, pp. 483–492, 2008.

[27] F. Kleber, T. John, and J. Harrington, "The implications for speech perception of incomplete neutralization of final devoicing in German," *Journal of Phonetics*, vol. 38, pp. 185–196, 2010.