How Are Word-final Schwas Different in the North and South of France?

Rena Nemoto¹, Martine Adda-Decker²,³

¹Institute of Cybernetics at Tallinn University of Technology, Tallinn, Estonia
²LPP CNRS/Université Paris 3, Paris, France
³LIMSI-CNRS, Orsay, France

Abstract

The aim of this paper is twofold: (i) give a large-scale description in realized word-final schwas of French lexical words for different regions (North vs. South) and different speaking styles (read vs. spontaneous speech); (ii) highlight differences in prosodic features and test these differences via automatic classification techniques. The proposed study relies on a subset of 12.5 hours of the French PFC corpus. Manually transcribed speech was segmented and labeled using automatic speech alignment and a pronunciation dictionary including optional word-final schwas for all words ending in a consonant. \( f_0 \) and intensity values were extracted and averaged over segments. Our study revealed that, for both speaking styles, word-final schwas of southern French tended to keep relatively high \( f_0 \) values and longer durations in comparison with northern French where \( f_0 \) tends to drop on a word-final schwa. On average, spontaneous speech featured smaller \( f_0 \) drops between final full vowel and subsequent word-final schwa vowel as well as longer durations. The automatic North/South classification of word-final schwas achieved better results for spontaneous speech. As for distinguishing between speaking styles, southern French obtained slightly better scores than the northern varieties.

Index Terms: French, word-final schwa, prosody

1. Introduction

The factors favoring the realization of the French mute “e” or schwa have been studied for decades by many authors. For example, the word \textit{vase} can be pronounced either /vaz/ or /vazə/. When realized, the schwa vowel has been found to be close to a mid-open or mid-closed somewhat rounded vowel /œ/ or /ø/ [1]. Similarly, the question whether and where the schwa is realized or elided has been largely investigated with comparative studies between northern and southern French, the northern variety being generally considered as “standard” French. The phonological representations of schwa in both varieties of French have also been extensively discussed [2, 3, 4, 5]. Southern French distinguishes itself from northern French by several characteristics, among which a higher production of schwas [2, 3, 4]. In particular, word-final schwas are typically omitted in standard French, whereas speakers tend to pronounce them rather systematically in the southern variety. In this contribution, our focus goes to word-final schwas. However, our main interest is not to determine whether schwas are present, absent, or just gradiently present. The question, which gave rise to this study, is the following: do word-final schwas, when realized in the North, have similar properties as southern word-final schwas? In particular, how do their prosodic properties compare? Some studies suggest that the fundamental frequency of a word-final schwa, which tends to drop after a full (tonic) syllable, is not necessarily falling in southern French [6]. How do they change with speaking style: read speech vs. spontaneous speech? Read speech is a somewhat unnatural way of speech production which may evoke memories of earlier school days when normative writing and speaking was a major goal to strive for. Our hypothesis is then that, if there are differences between North and South, they should appear more clearly in spontaneous speech than in the more normative read productions.

This study focuses on the prosodic realization of schwa vowels in northern and southern French. It is based on the investigation of a large corpus of spoken French, enabled by automatic speech processing. Section 2 gives an overview of the corpus and the adopted methodology. In Section 3, we describe the prosodic profiles of the word-final schwas across two regions (North and South) and two speaking styles (read and spontaneous speech). Section 4 shows the automatic North/South classification using prosodic features as well as the read/spontaneous speech classification. Conclusions and future work are presented in Section 5.

2. Corpus and methodology

2.1. Corpus

We used parts of the corpus collected in the framework of “Phonology of Contemporary French (Phonologie du Français contemporain)” (PFC) project [7]. Our subset consists of only male speakers with two different speaking styles: (1) free and guided face-to-face conversations as spontaneous speech; and (2) read text as read speaking style. For this study, we took into account 11 investigation points, divided into two regional categories: North and South. North contains a total of 32 speakers from 6 points in France (Aveyron-Paris, Brunoy, Dijon, Lyon-Villeurbanne, Roanne, Vendée) and one in Switzerland (Nyon). South includes 18 speakers from 4 points in the South of France (Biarritz, Douzens, Lacaune, Rodez).

2.2. Methodology

The manually transcribed speech data were phonemically and lexically labeled by automatic alignment [8] with a pronunciation dictionary proposing optional word-final schwas for all words ending in a phonemic consonant. Fundamental frequency \( f_0 \) and intensity values were measured using PRAAT [9]. Measurements were taken every 5 ms and averaged over segments. Durations were taken from automatic alignment.

In this study, we only focused on lexical words, grammatical words were excluded. Table 1 gives the number of occurrences of mono- to trisyllabic lexical words ending in a vowel (top), in a consonant without word-final schwa (middle), end-
Table 1: Number of lexical words of 1-3 syllables, ending in V, C and Ca as well as word-final schwa realization rates (%Ca#).

<table>
<thead>
<tr>
<th></th>
<th>Text North</th>
<th>Text South</th>
<th>Conversation North</th>
<th>Conversation South</th>
</tr>
</thead>
<tbody>
<tr>
<td>V#</td>
<td>24,689</td>
<td>10,781</td>
<td>3,579</td>
<td>2,009</td>
</tr>
<tr>
<td>C#</td>
<td>11,985</td>
<td>5,323</td>
<td>3,102</td>
<td>1,489</td>
</tr>
<tr>
<td>Ca#</td>
<td>972</td>
<td>1,069</td>
<td>468</td>
<td>580</td>
</tr>
<tr>
<td>(%Ca#)</td>
<td>(7.5)</td>
<td>(16.7)</td>
<td>(13.1)</td>
<td>(28.0)</td>
</tr>
</tbody>
</table>

computed with respect to the total of C-final words (C#+Ca#). As observed in the literature [3], the South of France has more realized word-final schwas than the northern group. Interestingly, spontaneous speech features about twice as many realized word-final schwas than read speech and southern speakers twice as many as northern speakers.

### 3. Prosodic investigation

Figure 1 shows a spectrogram of the French word domaine uttered by a speaker from the South (Douzens) with word-final schwa [ðɔmʁã]. The fundamental frequency ($f_0$), added as blue line, shows that $f_0$ is not falling on the word-final schwa as it tends to do on northern schwas. This phenomenon has already been mentioned by other authors [10]. However, a large-scale validation of this southern French specificity is still missing to the best of our knowledge. To this purpose, we first compute average prosodic profiles with respect to word syllabic length. Details of profile computations can be found in [11]. In the adopted notation $n_s$, $n$ gives the syllabic length of the investigated word subset and $s$ is a toggle indicating absence/presence of word-final schwa. Here, $s = 1$ indicates that a word-final schwa was realized (according to automatic alignment). In all presented figures, profiles are right-justified on $n$.

#### 3.1. Fundamental frequency ($f_0$)

The fundamental frequency ($f_0$) in Hz was converted into semitones (st) with 120 Hz as baseline frequency. Figure 2 (top) shows the average $f_0$ profiles of lexical words for read speech comparing North (left) and South (right) in terms of different syllabic word lengths ($n=1$-3). At a glance, these two figures are quite similar. We can observe common points for both regions as follows: (1) $f_0$ rise on the final syllable $n$; (2) $f_0$ falling from the final syllable $n$ to the word-final schwa. Whereas average $f_0$ profiles drop from the final (non-schwa) syllable $n$ to the word-final schwa, these drops seem to be slightly weaker for the South. However, as one may predict for a normative speaking style, read speech does not feature important differences between regions. $f_0$ profiles of conversational speech, shown in Figure 2 (bottom), reveal interesting differences. In northern speech (left), $f_0$ drops on word-final schwas, whereas $f_0$ profiles remain quite flat for word-final schwas in southern data, which tends to strengthen differences in word-final schwa realizations between North and South. With respect to speaking style, we may summarize as follows: for North (left of Figure 2, top: read text; bottom: conversation), an $f_0$ drop is observed on word-final schwa in both speaking styles with a slightly smaller $f_0$ drop for conversational speech. As for South (right of Figure 2), we can notice that profiles behave differently: the small final $f_0$ drop for read text (top) changes to a high stable plateau for spontaneous speech (bottom). To verify whether these $f_0$ differences between North and South and styles are significant, we conducted statistical tests (Wilcoxon test using the R software [12]). The differences between final full vowel $n$ and word-final schwa were statistically significant for both speaking styles ($p<0.0001$) and regions ($p<0.0005$).

#### 3.2. Duration

Figure 3 (top) shows the mean duration profiles of lexical words with word-final schwa in read text for the North (left) and South (right). The profiles of two figures are fairly similar. As observed in $f_0$, long final full vowel $n$ and shorter word-final schwa are measured. These long duration and high fundamental frequency ($f_0$) at the position $n$ are very common in French. As for $f_0$, differences between final full vowel and word-final schwa are smaller in the South. Duration profiles of conversational speaking style are given in Figure 3 (bottom). Similar to $f_0$ profiles, a duration drop is observed on word-final schwa in the North (left). Profiles tend to be stable from final full vowel to word-final schwa for South. As for read speech in Figure 3 (top), final full vowels $n$ are long in both regions. Next, we compared different speaking styles per region. For North (left of Figure 3, read text: top and conversations: bottom), low word-
Figure 3: Duration profiles of mono- to trisyllabic lexical words with additional word-final schwas. North (left); South (right). Read text (top); spontaneous conversations (bottom).

Statistical tests concerning word-final schwa duration were conducted both on absolute schwa duration and on intervocalic duration differences with respect to the final full vowel n duration. Using the Wilcoxon test, durations of word-final schwa in North and South are significantly different (p<0.0001) for read text. However, no significance can be shown in conversational speaking style. Next, differences in intervocalic duration deltas (the duration between a center of schwa duration and a center of previous final full vowel n) between North and South are statistically significant for conversations (p<0.0001), but no statistical significance could be found in read text.

With respect to speaking styles, the Wilcoxon test shows significant differences in word-final schwa durations (p<0.0001) in the North. However, any significance can be shown in the South. By contrast, intervocalic duration deltas fail to give any statistical significance in the North, whereas this measure is highly significant in the South (p<0.0001).

3.2.1. Duration ratio

Given the above observations, we introduced duration ratios between word-final schwa and its preceding final full vowel:

\[
\text{duration ratio} = \frac{\text{final schwa duration}}{\text{final full vowel duration}}
\]

A duration ratio above 1 implies that word-final schwa is longer than the final full vowel duration. Table 2 gives duration ratios of lexical words per region (North vs. South) and per speaking style (read vs. spontaneous speech) with mean, standard deviation (sd), and median values. For both read text and conversational speech, duration ratios are higher in the South than in the North. For both regions, conversational speech has higher mean (above 1) and median ratios as compared to read speech. As before, the same set of statistical tests (Wilcoxon test) were conducted but here with the duration ratio measurements. Results revealed significant differences for all test combinations (p<0.001 for all tests). This suggests that the duration ratio can be viewed as an interesting cue for discriminating between regions as well as between speaking styles.

3.3. Intensity

Figure 4 (top) shows the mean intensity profiles of lexical words ending in a word-final schwa for read speech as produced by northern (left) and southern (right) speakers. The profile shapes of mono- and disyllabic words are similar for both regions with falling values from final full vowel n to word-final schwa position. Trisyllabic word profiles in the North have a fairly standard shape similar to the one of trisyllabic t0n, whereas southern trisyllabic words feature an interesting peak in word-initial position. As a general comment, we may notice a rather large intensity drop between word-initial vowel and word-final schwa in the South, the amplitude of which remains weaker in the North. This southern specificity may be further investigated in future studies. Intensity profiles of conversational speaking style are presented in Figure 4 (bottom). In both North (left) and South (right), intensity drops in word-final schwas. Concerning speaking style differences per region, for North (left of Figure 4, top: read text and bottom: conversation), intensity drops on word-final schwas in both speaking styles. Figures also suggest that intensity drops between final full vowel and word-final schwa are smaller for conversations than for read speech. The same observations globally hold for our southern speakers (right part of Figure 4, top: read text and bottom: conversation). As for f0 and duration, statistical tests (Wilcoxon) were carried out. Statistically significant differences between North and South are observed in read speech (p<0.0001), however no sig-

<table>
<thead>
<tr>
<th>Duration ratio</th>
<th>Read text</th>
<th>Conversation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>North</td>
<td>South</td>
</tr>
<tr>
<td>mean</td>
<td>0.78</td>
<td>1.06</td>
</tr>
<tr>
<td>sd</td>
<td>0.67</td>
<td>1.14</td>
</tr>
<tr>
<td>median</td>
<td>0.63</td>
<td>0.71</td>
</tr>
<tr>
<td></td>
<td>North</td>
<td>South</td>
</tr>
<tr>
<td>mean</td>
<td>0.86</td>
<td>1.20</td>
</tr>
<tr>
<td>sd</td>
<td>0.70</td>
<td>0.86</td>
</tr>
<tr>
<td>median</td>
<td>0.71</td>
<td>0.86</td>
</tr>
</tbody>
</table>
significant difference can be found in conversational speech. The intensity value deltas between final full vowel $n$ and word-final schwa were also tested using a T-test since these values had normal distributions. Results showed statistically significant differences in speaking styles for both regions (North: $p < 0.0005$; South: $p = 0.0001$). However, no test succeeded in showing statistically significant differences for all conditions.

4. Word-final schwa classification

In the preceding section, it was statistically shown that some prosodic features of word-final schwas might be relevant cues for discriminating between North and South and/or between read and spontaneous speech. Can these features effectively contribute to differentiate between regions and/or speaking styles? In this last section, we made use of our prosodic features to test their efficiency in an automatic classification task, similar to work proposed in [13] where schwa realisation rates were used among other features to characterize northern vs. southern French speech.

In the following, eight features were defined and tested for automatic classification using the data mining software Weka [14]. Four features concern segmental values: mean $f_0$, phoneme duration, mean intensity, voicing ratio (being defined as the number of voiced ($f_0 > 0$) frames divided by the total number of frames in the segment). Four more features describe the relation between word-final schwa and final full vowel: $\Delta f_0$, $\Delta$duration, $\Delta$intensity, duration ratio.

4.1. Classification experiments

Two different classification tasks were conducted considering two regions (North and South) in the first task and different speaking styles (read and spontaneous speech) in the second task. The classification experiments were run using 10-fold cross-validation and 5 different algorithms of Weka were tested (Multilayer Perceptrons, Support Vector Machines, C4.5 Decision Tree, Logistic Regression Models, and Random Forest).

Table 3: Rates of correct word-final schwa classification. Left: North vs. South; right: read vs. spontaneous speech.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>North vs. South</th>
<th>Read text vs. Conv.</th>
<th>North</th>
<th>South</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Read text</td>
<td>Conv.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MLP</td>
<td>64</td>
<td>69</td>
<td>70</td>
<td>71</td>
</tr>
<tr>
<td>SVM</td>
<td>62</td>
<td>62</td>
<td>68</td>
<td>65</td>
</tr>
<tr>
<td>C4.5</td>
<td>63</td>
<td>67</td>
<td>67</td>
<td>71</td>
</tr>
<tr>
<td>LRM</td>
<td>63</td>
<td>68</td>
<td>70</td>
<td>72</td>
</tr>
<tr>
<td>Random Forest</td>
<td>62</td>
<td>69</td>
<td>71</td>
<td>73</td>
</tr>
<tr>
<td>Mean</td>
<td>63</td>
<td>67</td>
<td>69</td>
<td>70</td>
</tr>
</tbody>
</table>

Table 3 gives the rates of correctly classified word-final schwas either into North and South regions for read and conversational speech (left part), or into speaking styles (read vs. spontaneous speech) for both North and South word-final schwas (right part). Concerning regional classification (left), we can notice that conversational speech provides better cues to discriminate between North and South than read text. Conversational speech obtained 67% of correct identification on average whereas read speech got 63%. These results are in line with our prosodic descriptions. It also confirms our initial hypothesis that read speech is a more normative speaking style with smaller regional differences, as shaped through years of national education school practice. In particular, southern read speech featured $f_0$ drops from the final full vowel to the final schwa, which tends to imitate northern speech. Furthermore, read speech features shorter word-final schwa durations in the South, similar to northern measurements. These cues become more informative in conversational speech with flatter word-final schwa $f_0$ profiles in the South and duration ratios close to 1.

The rates of correct classification into speaking styles are presented at the right side of Table 3 for northern and southern word-final schwas. It is interesting to note that the speaking style classification task achieves better results in general than the regional classification task. We observed a slightly better discrimination between read text and conversational speech for southern speakers. These results are in line with our measurements, since conversational speech in the South kept almost the same values for both $f_0$ and duration from final full vowel to word-final schwa while these values decreased in read text during prosodic observations. However it was not expected that the difference of North and South results were so small with only 1% on average. The low performance of the SVM classifier on southern speech partly explains this result.

5. Conclusions

In this paper, we explored the question of how word-final schwas, when realized, differ in northern and southern French and how these differences manifest themselves in different speaking styles (read and spontaneous speech). The focus was put on prosodic features. This study includes more than 3000 word-final schwas of lexical words (grammatical words were excluded) extracted from over 12 hours of male speech of the PFC corpus. We provided a large-scale description including $f_0$, duration and intensity lexical profiles. Among the observed southern specificities, we may highlight the relatively high $f_0$ values (plateau) for word-final schwas in southern French spontaneous speech as well as close to 1 duration ratios (ratio between schwa vowel and final full vowel durations) for both speaking styles. Northern French tended to feature systematic $f_0$ drops on word-final schwas and smaller duration ratios in all conditions. Our results also confirmed the hypothesis that, if there are differences between North and South, they should appear more clearly in spontaneous speech than in the probably more normative read productions. Finally, we tested our prosodic features using automatic classification (five classifiers) and 8 segmental and intersegmental prosodic features. On average, the best regional classification rate (67%) was achieved for spontaneous speech and the highest classification results (70%) were obtained in the speaking style task with southern speech. In the future, the presented investigations can be adapted to word-internal schwas in order to check for regional and/or speaking style specificities. Furthermore, investigations beyond the lexical domain are foreseen, however they require appropriate annotations into larger units such as accent/intonation phrases.

6. Acknowledgements

This research was supported by the European Regional Development Fund through the Estonian Center of Excellence in Computer Science (EXCS) and the Estonian Ministry of Education and Research target-financed research theme No. 0140007s12 to the 1st author. It was also supported by the French ANR LIBEX EFL (ANR/CIG) Program and PFC-Cor project.
7. References


