Phrase-medial vowel devoicing in spontaneous French

Francisco Torreira[1,2], Mirjam Ernestus[1,2]

[1]Radboud Universiteit Nijmegen, The Netherlands
[2]Max Planck Institute for Psycholinguistics, The Netherlands

`Francisco.Torreira@mpi.nl, Mirjam.Ernestus@mpi.nl`

Abstract

This study investigates phrase-medial vowel devoicing in European French (e.g. /ty pø/ [typø̥] 'you can'). Our spontaneous speech data confirm that French phrase-medial devoicing is a frequent phenomenon affecting high vowels preceded by voiceless consonants. We also found that devoicing is more frequent in temporally reduced and coarticulated vowels. Complete and partial devoicing were conditioned by the same variables (speech rate, consonant type and distance from the end of the AP). Given these results, we propose that phrase-medial vowel devoicing in French arises mainly from the temporal compression of vocalic gestures and the aerodynamic conditions imposed by high vowels.

Index Terms: vowel devoicing, French, spontaneous speech

1. Introduction

1.1. Vowel devoicing

Vowel devoicing occurs in many languages across the world, such as Japanese [1, 2, 3], Korean [4, 5], Greek [6], Turkish [7] and Québec French [8]. In this study, we investigate the occurrence and characteristics of phrase-medial vowel devoicing in European French, a language for which only phrase-final devoicing has been studied in the past.

Vowel devoicing may originate from different mechanisms, but it is not always clear which mechanism is responsible for vowel devoicing in a particular language. First, it can be the result of a high-level phonological process or rule by which the voicing feature of a vowel is modified to match that of an adjacent voiceless consonant. In support of this mechanism, [1] found that the vocal folds remain abducted during the production of Japanese devoiced vowels.

Vowel devoicing can also be regarded as a lower-level phonetic effect. For instance, [7, 4] have argued that vowel devoicing in Turkish and in Korean results from the overlap of the glottal gestures of vowels with those of their adjacent voiceless consonants. Along these lines, [2] observed that even if the glottis remains open throughout the production of voiceless vowels in Japanese, muscular activity is unusually variable during the production of these vowels, which suggests that the voicing gesture is somehow inhibited, but not completely deleted.

Another phonetic explanation involves the aerodynamics of vowel production. If the transglottal pressure differential necessary for voicing is not achieved, for instance during the production of an extremely constricted and short high vowel, voicing will not be possible even if the vocal folds are properly adducted. The cross-linguistic observation that vowel devoicing typically affects unstressed high vowels lends support to this aerodynamic explanation in many cases.

1.2. Vowel devoicing in French

European French is well known for its use of vowel devoicing in phrase-final position as a prosodic and discourse marker [9, 10]. This kind of devoicing affects high and mid vowels before a pause, and can be characterized as a long and sustained fricative noise preceded by a portion of a voiced vowel, rather than as a reduced or elided vowel.

On the other hand, phrase-medial devoicing has been only briefly reported for this language, although it appears to be recurrent in connected speech: in a study of short vowels (< 30 ms), automatically extracted from a corpus of conversational speech, [11] found that devoicing affected 20% of the high vowels examined, and in some rare cases, non-high vowels as well. In the vast majority of cases, the devoiced vowels were preceded, but not necessarily followed, by voiceless consonants.

1.3. The present study

In the present study we further document and investigate the occurrence of phrase-medial vowel devoicing in European French. By phrase-medial devoicing we understand devoicing in non-utterance-final position, as in the utterance /ty pø/: [typø̥] *tu peux* 'you can'. We first examine the frequency of complete and partial devoicing in a corpus of spontaneous casual French. Second, we investigate what factors condition the likelihood of complete devoicing and the extent of partial devoicing. Then we look for relationships between devoicing and syllable, stop closure and vowel durations. Finally, we examine if devoicing is accompanied by increased consonant-vowel (CV) coarticulation, as captured by spectral measures.

2. Method

2.1. Materials

Based on the observations made by [11], we decided to focus on vowels and on contexts for which phrase-medial devoicing is recurrent, that is on *unaccented high vowels preceded by a voiceless consonant* (/p/, /t/, /k/, /f/, /s/ or /ʃ/). In order to consistently measure the end of the target vowels, we decided to further restrict our study to vowels followed by a consonant involving a complete oral closure (i.e. a stop, either oral or nasal, and voiced or voiceless) in the same or following word. We considered both function and content words. On the basis of the restrictions above, only three function words were studied: *qui* /ki/, *si* /si/ and *tu* /ty/ 'you'. Since we restricted ourselves to unaccented function words, in our data the word *qui* was always a relative pronoun, and the word *si* was always a conditional conjunction ('if') or an adverb ('so'). As for content words, only words with target vowels in non-final position (where they cannot receive a phrasal accent) were considered. Tokens contain-

ing laughter, disfluencies or intrusive background noise were discarded. We also discarded tokens in which the postvocalic consonant did not exhibit a clear closure in the signal. Target vowels meeting the conditions above were randomly extracted from the Nijmegen Corpus of Casual Speech [12] until 250 content words (151 word types) and 300 function words (100 words for each function word type) were available for analysis.

2.2. Measurements and annotation

The first author made the following measurements and annotations for every consonant-vowel-consonant sequence (C_1VC_2) in the dataset:

- Closure duration: only for C_1 stops ($n = 192$).

- VOT: only for C_1 stops, from stop release to onset of voicing in V, or in its absence, to the start of the upcoming C_2 closure.

- Voicing duration: from the onset of voicing corresponding to the target vowel to the onset of the C_2 closure.

- Vowel duration: only for tokens with a stop as C_1, the sum of VOT and voicing duration, corresponding to the stretch of the signal in which the vocal tract was unconstricted between C_1 and C_2.

- Voiceless frication duration: only for C_1 fricatives, from increase in acoustic energy above 4 KHz to onset of voicing in the vowel, or in its absence, to the start of the following stop closure.

- Syllable duration: the sum of closure/voiceless frication, VOT and voicing duration.

- Prosodic context: presence of an Accentual Phrase (AP) or Intonational Phrase (IP) boundary before/after the target word. The annotation of AP and IP boundaries was carried out by the first author following [13].

- Number of syllables in the AP containing the test word, and distance in number of syllables from the target syllable to the AP start and end.

- Speech rate: (number of syllables in AP - 1) / (Duration of AP excluding last syllable). We excluded the last syllable in the AP from the computation of our speech rate estimate to avoid variability due to final lengthening.

A check was performed in order to assess the reliability of the measurements and annotations. Forty tokens from the dataset were randomly selected, and independently analyzed by a trained transcriber unaware of the purposes of our study. No major disagreement was found between our measurements and those of the independent transcriber (100% of closure duration, 91% of VOT and 96% of voicing duration measurements differed by 10 ms or less between the two annotators; the annotation of the location of preceding and following intonational boundaries coincided in 82% and 92% of the cases, and the annotation of the type of boundary in 92% and 77% of the cases).

2.3. Results

2.3.1. Frequency of complete devoicing and distribution of voicing duration

More than a third of the target vowels ($n = 199$, 36.3%) were completely devoiced (no voicing was found between the prevocalic and postvocalic consonants). The voicing durations of the remaining vowels formed a unimodal distribution with a mean of 30 ms and a standard deviation of 14 ms. A considerable

number of tokens had voicing durations below 15 ms ($n = 49$). In agreement with [11], these numbers indicate that vowel devoicing is a recurrent phenomenon in European French. Figures 1 and 2 illustrate cases of complete and partial devoicing in our data.
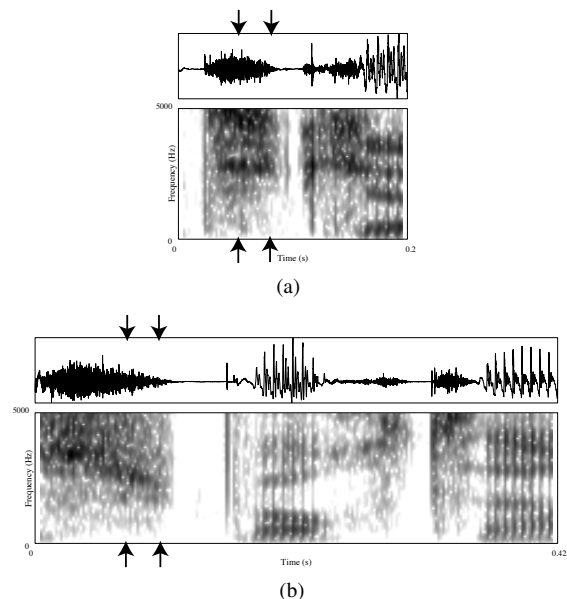

(a)

(b)

Figure 1: Examples of complete vowel devoicing in the words (a) *ticket* /tikɛ/ 'ticket' and (b) *supporter* /sypɔʁtɛʁ/ 'sports fan'. Arrows indicate the approximate location of the devoiced vowel.

2.3.2. Conditioning variables

We investigated which of the variables that we had annotated conditioned the occurrence of complete devoicing and the duration of vocalic voicing. If complete and partial devoicing are different points in a phonetic continuum, they should be conditioned by the same factors. Based on previous research, we expect devoicing to occur more often at faster speech rates and in non-prominent prosodic locations (far from phrasal prosodic edges, word-medially) [4, 7]. If devoicing is seen as a reduction phenomenon, it can also be expected to occur more often in function than in content words. We first fitted a logistic regression model with complete devoicing as the response and the following predictor variables: speech rate, vowel type (/i/, /y/ or /u/), C_1 manner, C_1 place of articulation, C_2 voicing, preceding and following prosodic context, AP length, distance to AP end, distance to AP start, position in the word (word-initial vs. word-medial syllable), word class (content vs. function word). Speaker was included in the model as a random factor.

Three variables were found to favor the probability of complete devoicing. First, complete devoicing was more probable the faster the speech rate ($\beta = -0.16, z = -3.53, p < .0005$). Second, devoicing tended to occur more often after fricatives than after stops ($\beta = -0.52, z = -2.65, p < .01$). Third, devoicing was more probable the further away the target vowel was from the right edge of its AP ($\beta = -0.63, z = -4.15, p < .0001$). No other variable was a significant predictor in the model.
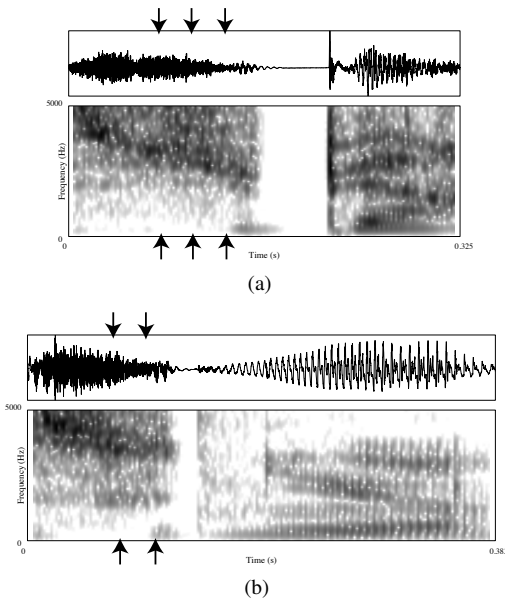
We then fitted a linear regression model using only tokens

(a)

(b)

Figure 2: Examples of partial vowel devoicing in the words (a) *super* /sypɛʁ/ 'super' and (b) *soutenue* /sutny/ 'formal'. Arrows indicate the approximate location of the devoiced vowel.

with positive voicing durations. This model had voicing duration as the response, and the same variables listed above as predictors. The obtained model was similar to the logistic regression model above, except for a significant interaction between C_1 manner and word class. Voicing duration was shorter at faster speech rates ($\beta = -1.07, t = -2.99, p < .005$) and the further away the target vowel was from the end of the AP ($\beta = -4.15, t = -3.096, p < .005$). As for the interaction, voicing duration tended to be shorter after fricatives than after stops ($\beta = -4.4, t = -2.07, p < .05$), but this effect did not appear to hold for function words. Since the only function word having a fricative was *si*, it cannot be concluded if the interaction should be attributed to the word *si*, to consonant /s/ or to fricatives in function words.

We conclude from these analyses that complete and partial devoicing were generally sensitive to the same factors, namely speech rate, C_1 manner and the distance from the target syllable to the end of the AP. This suggests that phrase-medial vowel devoicing in European French is a continuous process, with completely devoiced vowels at its endpoint.

2.3.3. Relationships between devoicing and syllable, closure and vowel durations

Previous research on several languages has shown that vowel devoicing is usually accompanied by a temporal compression of the vowel and its surrounding consonants [7, 4]. We checked if this is also true for French by looking for statistical relationships between the occurrence of complete devoicing and syllable, closure and vowel durations.

A logistic regression model with complete devoicing as response and syllable duration as predictor indicated that complete devoicing was more likely the shorter the target syllable ($\beta = -0.026, z = -7.72, p < .0001$). In order to control for potential confounds, we fitted a second model in which syllable duration had been normalized with respect to its segmental

composition, speech rate and its distance to the end of the AP. This was done by using as predictor the residuals of a linear model predicting syllable duration with these variables. Again, we found that devoicing was more likely the shorter the syllable ($\beta = -0.022, z = -6.31, p < .0001$). These results are illustrated in Figure 3, which shows syllable duration as a function of devoicing. This finding suggests that phrase-medial devoicing in French involves some sort of temporal compression at the supraglottal level.
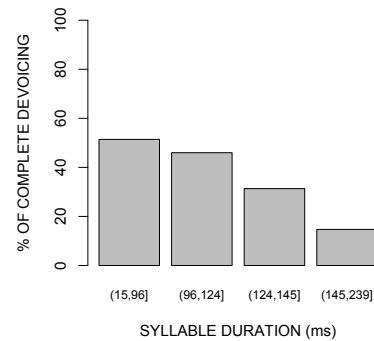

Figure 3: Percentage of complete devoicing as a function of syllable duration for four quantiles of equal probability (.25).

We then investigated if this temporal compression equally affected the consonantal and vocalic parts of the syllable. Because the consonantal and vocalic parts in a syllable with a fricative onset and a devoiced vowel cannot be measured reliably on the basis of a discrete acoustic landmark (e.g. a stop release), we only considered syllables with intervocalic stop onsets. We fitted regression models with closure and vowel duration as responses, the occurrence of complete devoicing as the main predictor and speech rate, C_1 and V types, and distance from the target syllable to the end of the AP as covariates. The model predicting closure duration showed that this variable was not affected by devoicing ($\beta = 3.04, t = 1.27, p > .2$). In other words, the duration of stop closures did not appear to differ between syllables with a devoiced vowel and syllables with a voiced vowel. On the other hand, we found a significant effect of devoicing on vowel duration (defined here as the interval from the C_1 stop release to the C_2 closure onset, see above). Devoiced vowels tended to be significantly shorter than voiced vowels ($\beta = -28.62, t = -28.37, p < .0001$). It appears therefore that devoicing is accompanied by the temporal reduction of only the oral part of the syllable, rather than of the syllable as a whole. This finding is illustrated by Figure 5, which shows the percentage of complete devoicing as a function of vowel duration.

2.3.4. Complete devoicing and the spectral characteristics of C_1

It has been claimed for Japanese, Turkish and Korean that vowel devoicing occurs as the result of increased gestural overlap between vocalic and consonantal gestures [3, 7, 4]. We now investigate if this is true for French. Under this hypothesis, it can be expected that devoiced vowels will exhibit signs of increased CV coarticulation. Since the acoustic effects of coarticulation depend greatly on the specific consonant and vowel sequence
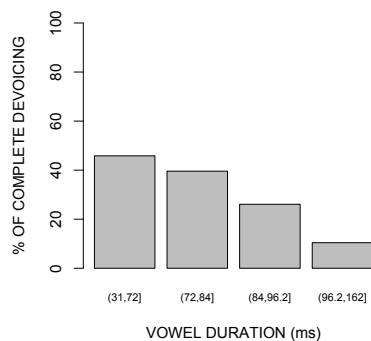
Figure 4: Percentage of complete devoicing as a function of vowel duration for quantiles of equal probability (.25).

under examination, we focused only on sequences for which at least 100 tokens were available for analysis: /ki/, /ty/ and /si/. The vast majority of tokens were instances of the function words *qui*, *tu* and *si*. In the case of /ki/ and /ty/ we calculated the spectral center of gravity and the spectral peak in a 20 ms Hamming window left-aligned with the stop release. In the case of /si/, we calculated the spectral center of gravity 25% into the voiceless frication interval (measured from the left) in a 0.5-10 KHz band. In all three syllables, we examined if devoicing had an effect on these spectral measures. In the case of /ki/, increased coarticulation between /k/ and /i/ should translate into increased /k/ fronting and higher values for the two spectral parameters considered. On the other hand, in the case of /ty/ and /si/, increased palatalization of /t/ and /s/ should result into a downward shift of these parameters. In all models, gender was included as a covariate, and speaker as a random factor.

For /ki/ and /ty/, we did not find any statistical spectral differences between tokens with devoiced and voiced vowels. However, in all four models the signs of the obtained regression coefficients were consistent with the hypothesis of increased coarticulation for syllables with devoiced vowels (/ki/: $\beta = 90.5$ and $\beta = 213.4$; /ty/: $\beta = -152.6$ and $\beta = -180.4$). For /si/, we found a significant difference in the direction predicted by the hypothesis ($\beta = -344.5, t = -2.13, p < .05$). It should be noted that the number of available tokens was considerably higher for /si/ than for /ki/ and /ty/ (/si/: $n = 145$, /ki/: $n = 104$, /ty/: $n = 115$). For this reason, it can be speculated that statistically significant differences for /ki/ and /ty/ might have been observed if we had used bigger samples. In summary, we found some acoustic evidence of increased CV coarticulation in cases of complete devoicing.

3. Discussion

We have shown that phrase-medial devoicing occurs frequently in connected European French. In our data, over a third of the studied high vowels are completely devoiced, and an important number of tokens are partially devoiced in a considerable degree, both in function and content words. Devoicing occurs for syllables of all durations, but it tends to be more frequent in syllables with temporally reduced vowels. Study of the spectral characteristics of C_1 provided some evidence of increased coarticulation between C_1 and V in cases of complete devoicing, especially for syllables with fricative onsets. Finally, we found

that the occurrence of complete and partial devoicing are conditioned by similar variables (speech rate, manner of articulation and distance to the upcoming AP boundary). Surprisingly, we did not find differences in the likelihood and extent of devoicing between function and content words.

In view of these results, we propose that phrase-medial vowel devoicing in European French is a phonetic phenomenon arising mainly from temporal compression and affecting a class of vowels which, by their constricted nature, are unfavorable to the production of voicing for aerodynamic reasons. Although the observed increased CV coarticulation in devoicing cases suggests that articulatory overlap between C and V may be the main cause of this vowel shortening, a simpler mechanism of truncation of the vocalic gestures should be considered as well. Articulatory data are needed to elucidate this question.

Finally, given the high frequency of phrase-medial devoicing in French, one may speculate to what extent it has been or is being conventionalized as a proper phonetic target, or whether, on the other hand, it is merely the byproduct of the mechanical and phonetic factors mentioned above. In any case, it appears that the observed pattern of vowel devoicing in French may lay the foundation for future phonologizations, which, for instance, might lead to the emergence of a class of voiceless vowels or affricated consonant allophones in this language.

4. References

[1] Hirose, H. (1971). The activity of the adductor laryngeal muscles in respect to vowel devoicing in japanese. *Phonetica*, 23:156–170.

[2] Yoshioka, H. (1981). Laryngeal adjustments in the production of the fricative consonants and devoiced vowels in Japanese. *Phonetica*, 38:236–251.

[3] Beckman, M. E. and Shoji, A. (1984). Spectral and perceptual evidence for CV coarticulation in devoiced /si/ and /syu/ in Japanese. *Phonetica*, 41:61–71.

[4] Sun-Ah Jun, Mary Beckman, H.-J. L. (1998). Fiberscopic evidence for the influence on vowel devoicing of the glottal configurations for korean obstruents. *UCLA Working Papers in Phonetics*, 96:43–66.

[5] Mo, Y. (2007). Temporal, spectral evidence of devoiced vowels in Korean. In *Proceedings of ICPhS 2007, Saarbrücken*.

[6] Dauer, R. M. (1980). The reduction of unstressed high vowels in modern Greek. *Journal of the International Phonetic Association*, 10:17–27.

[7] Jannedy, S. (1995). Gestural phasing as an explanation for vowel devoicing in Turkish. *Ohio State University Working Papers in Linguistics*, 45:56–84.

[8] Cedergren, H. . L. S. (1985). La chute des voyelles hautes en français de Montréal: 'As-tu entendu la belle syncope?'. In Lemieux, M. and Cedergren, H., editors, *Les tendences dynamiques du français parlé à Montréal*, pages 57–145. Bibliothèque nationale du Québec.

[9] Fagyal, Z. and Moisset, C. (1999). Sound change and articulatory release: where and why are high vowels devoiced in Parisian French. In *Proc. of the 14th ICPhS*, pages 309–312, San Francisco.

[10] Smith, C. L. (2003). Vowel devoicing in contemporary French. *Journal of French Language Studies*, 13(2):177–194.

[11] Meunier, C., Meynadier, Y., and Espesser, R. (2008). Voyelles brèves en parole conversationnelle. In *Actes, Journées d'Etude sur la Parole 2008*, pages 97–100.

[12] Torreira, F., Adda-Decker, M., and Ernestus, M. (2010). The Nijmegen Corpus of Casual French. *Speech Communication*, 52:201–121.

[13] Fougeron, C. and Jun, S.-A. (1998). Rate effects on French intonation: Prosodic organization and phonetic realization. *Journal of Phonetics*, 26:45–69.