Phonetic reduction versus phonological deletion of French schwa: Some methodological issues

Audrey Bürki a, Cécile Fougeron b,*, Cedric Gendrot b, Ulrich H. Frauenfelder a,c

Abstract

A categorical phonological process of deletion is traditionally assumed to account for the alternation of schwa with zero in French. This process is assumed to result in two discrete outputs: forms with schwa (i.e., schwa variants) and forms without schwa (i.e., non-schwa variants). However, the two studies we present here suggest a more complex picture. In the first study, we investigate the phonetic variability of schwa in a large number of occurrences of schwa variants and find that schwa, like other segments in French, undergoes phonetic reduction. As a consequence, some tokens without schwa in connected speech may be the result of a process of gradual phonetic reduction rather than the result of a categorical process of alternation. In the second study, we examine the perception of schwa word tokens extracted from connected speech. We show that deciding whether a token was produced with or without the schwa is not always possible. Furthermore, listeners rely on other types of cues than acoustic ones in order to make their judgements (i.e., speech rate, word length and segmental context). These findings have important theoretical and methodological implications that must be taken into account in the empirical study of French schwa alternation.

1. Introduction

Over the years, studies on continuous speech in various languages have revealed a great amount of variation in the pronunciation of words with respect to their assumed canonical form. Along with changes in segmental properties, the disappearance of segments in the speech signal is frequent (e.g., Ernestus, 2000 for Dutch; Helgason & Kohler, 1996 for German; Johnson, 2004 for American English). The absence of an expected segment in the realization of a word in running speech is usually accounted for by a gradual process of phonetic reduction. A major argument for this interpretation is found in the variability of the observed outputs; acoustic surface forms present a continuum of realizations ranging from temporally and spectrally reduced forms to forms with subtle remnant acoustic cues, or forms with total disappearance of cues (Davidson, 2006; Helgason & Kohler, 1996; Manuel et al., 1992). Gradient reduction processes have been documented in French for vowels other than schwa and for consonants (see for instance Su, 2003).

In this language, such phonetic reduction coexists with another process, segmental “deletion”, the alternation of the vowel schwa with zero in some French words. For example, the word fenêtre ‘window’ has an alternating pronunciation between a schwa variant [fɛntɛʁ] and a non-schwa variant [fɛntɛʁ]. This alternation has received a lot of attention in the linguistic literature, and much debate has taken place between different theoretical accounts: schwa deletion rule, schwa insertion rule, allophony or allomorphy (see Côté & Morrison, 2007 for a review). These accounts all agree that schwa-zero alternation in French is not a phonetic process. Schwa and non-schwa variants are considered to be categorically distinct, and the articulatory target for schwa is assumed to be either present or absent depending on the output variant.

As a consequence of this conception of French schwa alternation as a mere phonological process, the phonetic variability of its realization is never mentioned in the literature. It is, however, reasonable to expect that, like other segments, schwa should undergo temporal and spectral reduction in running speech. It should thus present gradient realizations ranging from weakly reduced to extremely reduced forms, and this phonetic reduction could possibly result in the disappearance of vocalic cues in the acoustic signal. We address this issue in our first empirical study, whose aim is to document the phonetic (temporal and spectral) variability of schwa realization in running speech and estimate the proportion of strongly reduced schwas. We do so by means of an acoustic study involving over 3000 occurrences of schwa words in...
a radio broadcasted corpus. If we assume that there is indeed a
categorical schwa-zero alternation and find that, as expected,
schwa is also subject to phonetic reduction, some methodological
issues arise about what should be considered as a schwa or a non-
schwa variant. These concerns are directly linked to the general
topic of this special issue of the Journal of Phonetics and relate to
the distinction between (extreme) phonetic reduction and the
categorical deletion of segments.

If, as predicted, acoustically reduced forms of schwa occur in
running speech, deciding whether an occurrence does or does not
contain a schwa may not always be as straightforward as it is
generally assumed. Many studies on segmental reduction in
different languages have acknowledged that the presence/absence
of a given segment in connected speech may be difficult to
determine for some tokens (e.g., Pitt, Johnson, Hume, Kiesling,
and Raymond, 2005) and that disagreement between raters exists.

In the case of French schwa, the existence of ambiguous forms is
(almost) never mentioned. Only one author (Cornulier, 1975)
mentioned the existence of numerous ambiguous tokens. In the
second empirical study of this paper, we question the fact that
the perceptual discrimination of forms produced in continuous
speech as being a schwa or a non-schwa variant is as trivial as is
usually assumed. In addition, we examine whether listeners rely
on cues other than acoustic information to categorize the forms. If
acoustical cues are indeed insufficient to categorize some
occurrences, it is likely that listeners will rely on other types of
information (e.g., variant frequency, segmental context or speech
rate). We address these questions by first screening over 4000
tokens extracted from a corpus of connected speech and then
performing a perception test with 22 participants. If we indeed
find that tokens of schwa words in connected speech are not
always unambiguously classified as schwa versus non-schwa
variants, methodological questions arise concerning the criteria
and methods used to categorize output forms.

Before turning to our studies, we will discuss some preliminaries
regarding schwa-zero alternation in French. In French, it is agreed
that the term “schwa” does not cover a unified entity. In the present
paper, we are concerned with alternating schwas (also called
‘optional schwas’) occurring word internally. Alternating schwas
occur in words whose pronunciation alternates between a schwa
and a non-schwa variant (e.g., semaine ‘week’ [semɛn] or [semɛn],
revenir ‘come back’ [ʁəvœ̃ni], [ʁavœ̃ni] or [ʁavœ̃ni]). The inventory
of alternating words is difficult to define. Dialectal, ideological, and
sociolectal factors affect whether a word alternates between a
schwa and a non-schwa variant. For example, most Swiss French
speakers will refer to Geneva as [ʒɛvɛ], while most French speakers
will always use the schwa variant [ʒœ̃v]. In standard French,
speakers will tend to use both variants for many words with a
schwa in the first syllable (i.e., initial schwas) but only the non-
schwa variant for words with a schwa in a medial syllable (e.g.,
casserole ‘pot’ [kaskœ̃l]). By contrast, in the region of Aix-Marseille
(southern French), most words with initial schwas tend to be
produced in their schwa variant only, whereas words with medial
schwas will often be produced in their two variants.

2. Speech material

The speech material used in the two studies presented here
comes from a subset of the ESTER corpus (Galliano et al., 2005). It
contains 24 h of radio broadcasted news including the produc-
tions of a total of 574 speakers. In order to select words with an
alternating schwa, a repertory of French words containing a
schwa vow[1] was constituted by merging several French lexical
databases (see Burki, Gendrot, Gravier, Linares, & Fougeron, 2008
for further details). Since the definition of schwas as either
alternating or non-alternating varies according to the databases,
all vowels labelled as schwas (either optional or obligatory) were
included at this stage. A first filter was applied to eliminate
schwas in compounds, at word and clitic boundaries, before
derivational suffixes or inflectional endings. The ESTER corpus
was then searched for spoken occurrences of these words and
sound files were extracted. In order to be able to compare a given
word across speakers and productions, we restricted the dataset
to the words whose tokens appeared at least 5 times in the
corpus. This led us to a total of 22,773 occurrences corresponding
to 583 different words. The tokens had then to be classified as
alternating versus non-alternating. We used an objective method,
based on an analysis of the realizations of these words (always
produced with schwa or always produced without schwa versus
alternating pronunciations) in the recordings. Speakers’ produc-
tions were screened with the IRISA (Institut de Recherche en
Informatique et Systèmes Aléatoires) automatic recognition
system, and words that were identified as having occurrences of
both schwa and non-schwa variants were selected. For example,
the word demi ‘half’ was recognized 13 times as [dɔmi] and 17
times as [dɛmi], and was thus selected. The items always produced
with the same pronunciation variant (schwa or non-schwa
variant) were classified as non-alternating, and thus disregarded.

The output of the system was then checked manually. A single
judge listened to all occurrences and eliminated clear recognition
errors (wrong lexical items), productions of non-native speakers,
and unintelligible/noisy speech files (totalizing 8.6% of the data).
Occurrences with an obviously missed or wrongly inserted schwa
were corrected (and double-checked by a second judge who
agreed on the errors for all items without exception). This
procedure led to a set of 4294 tokens of alternating words. These
tokens corresponded to 191 different lexical items produced by
361 speakers. There were 134 words with an initial schwa and 57
words with a word-medial schwa. Speakers’ dialects could for the
most part be classified as “standard French”. It must be noted
that, compared with a more spontaneous register, the number of
alternating schwas and of non-schwa variants (28% see below) is
likely to be reduced in a radio corpus where some productions are
pre-planned speech.

3. Study 1: Acoustic examination of the variability
of schwa variants

Our aim in this acoustic investigation is to describe the
phonetic variability of realized schwas in schwa variants. Both
durational and spectral properties of schwa are examined. We are
particularly interested in the distribution of schwa duration and
in the proportion of very short schwas. The spectral changes
associated with vowel duration are also of interest. Shortened
vowels have been shown to be more affected by the surrounding
segmental context (see Mooshammer & Geng, 2008 for a recent
review). Our goal is thus to determine whether schwas behave
like other vowels in this respect.

3.1. Method

The alignments provided by the automatic recognition system
for the 4294 schwa-zero alternating forms extracted from the
ESTER corpus were manually corrected. A trained phonetician

---

[1] Note that if schwa is commonly transcribed as “ə” in the literature, it is not
as “neutral” or central as the position of “ə” in the IPA chart suggests (Ayres-
Bennett & Carruthers, 2001). The exact spectral characteristics of schwa as well as
their similarities/differences with that of the vowels [ɛ] and [e] are still debated in
the literature (see e.g., Fougeron, Gendrot, & Burki, 2007).
undertook the verification of all the tokens, and 47% of these were
double-checked by another trained phonetician (with a high
degree of inter-judge reliability, see Bürki et al., 2008). The
recordings were analyzed based on spectrographic and waveform
displays, and the following criterion was taken to define the
absence/presence of schwa; a schwa label was given to portions of
the signal showing periodicity and formant structure, even if this
portion was very short. The onsets and offsets of the second
formant were used to define the boundaries of this vocalic
portion. This criterion for segmentation of the vowel intervals was
the best compromise found given the wide variety of consonantal
contexts. Potentially fully devoiced vowels or other phonetic
properties that could be attributable to schwa were not
considered for schwa labelling. On this basis, 3098 tokens out of
the 4294 (72%) were labelled as schwa variants. The duration of
the vocalic portions and spectral properties in terms of F1 and F2
frequencies (measured between 1/3 and 2/3 of the vowel and
reduced to an averaged value) for these 3098 tokens were
automatically computed with Praat (Boersma & Weenink, 2007).

3.2. Results

Based on the segmentation adopted (onset–offset of F2),
schwas are found to be on average 51 ms long (sd=18 ms,
median=50 ms), which corresponds to 14% of the total word
length. Schwas in initial syllables (initial schwas) had an average
duration of 51 ms (15% of word length), whereas schwas in a
medial syllable had an average duration of 44 ms (10% of word
length). Confirming previous studies (e.g., Adda-Decker, 2006;
Bartkova & Sorin, 1987), schwa is a rather short vowel overall.
This fact can in part be explained by positional restriction
according to which this vowel never occurs in an accented
syllable in French. In comparison with other French vowels
produced in the ESTER corpus, but analyzed automatically
(i.e., without manual correction), Fougeron et al. (2007) reported
a mean duration of 67 and 65 ms for the vowels /a/ and /æ/,
respectively (with schwa=50 ms), Gendrot and Adda-Decker
(in press) reported a mean duration of 70 ms for /æ/, 76 ms for
/j/ and 87 ms for /u/ in the same corpus.

As shown in Fig. 1, the distribution of schwa duration is close
to normal. Schwa duration ranges from 8 to 150 ms. Interestingly,
a non-negligible number of very short vowels can be observed at
the left-hand side of this distribution. One third of the schwas are
40 ms or shorter and 14% are 30 ms or shorter. According to
Meunier, Meynadier, and Espesser (2008) schwas below 30 ms can
be labelled "extra-short", and these extra-short schwas can be
considered as being temporally reduced vowels. In their corpus
of spontaneous French, they observed a proportion of 30%
extra-short vowels (including different vowels). The higher
proportion of extra-short vowels in their corpus is likely due to
speech style (casual conversation in their case, journalistic speech
in our corpus). Furthermore, in their study, devoiced vowels were
included, whereas we included only voiced segments.

In order to determine how the quality of schwa varies according
to its duration, we followed the procedure applied by Gendrot and
Adda-Decker (in press) to the other French vowels. Schwas of our
corpus were split into two categories (short: < 50 ms and long: ≥ 50
ms) so as to include a sufficient number of consonantal contexts in
each category. We then ran two linear mixed effects models, one
with F1 as the dependent variable, the second one with F2 as the
dependent variable. In both models, speaker and word were entered
as random terms, schwa duration (short versus long) and right and
left segmental context as predictors. In order to examine how
spectral variations are linked to preceding and following contexts,
these predictors were included in the analysis together with their
interaction with duration. Again following Gendrot and Adda-Decker
(in press), we defined four categories of contexts: labial, dento-
alveolar, palato-velar (including pre-palatal, palatal, and velar) and
uvular. The second formant (F2) is specifically included in our
analysis as it is known to vary greatly as a function of the place of
articulation of surrounding consonants (Delattre, Liberman, &
Cooper, 1955). If vowels do not reach their target when they get
shorter (Lindblom, 1963), then F2 mean values should get closer to

![Fig. 1. Distribution of the durations of the intervals with voicing and formant structure for the 3098 occurrences of schwa variants analyzed.](image-url)
the locus of the neighbouring consonants. $F_2$ mean values of schwas are thus expected to rise for short schwas compared to long schwas in contexts where the locus value of the consonant is high, that is, for alveolar, palato-velar, and uvular (locus at 1800, 3000 and 1400 Hz, respectively, Delattre et al., 1955). For the labial context, $F_2$ values of the schwa should not be affected by the labial consonant per se (as the tongue is not required for those consonants) but rather by the consonant on its other side, if not labial as well. In our corpus, in most cases (1292/1660), the schwas are in an asymmetric context with a labial on one side and a non-labial on the other side. In these cases, the raising of $F_2$ is due to the influence of the non-labial consonant (Gendrot & Adda-Decker, in press).

The value for $F_1$ is 409 Hz on average (405 Hz for short schwas and 412 for long schwas). The statistical model for $F_1$ shows no main effect of duration ($\beta = -3.7$, $t = -0.93$, n.s.) but an effect of right ($F(3, 2935) = 7.5$, $p < 0.0001$) and left ($F(3, 2935) = 35.7$, $p < 0.0001$) consonantal contexts. $F_1$ is higher if the preceding or the following consonant is uvular, as shown in previous research (e.g., Delattre et al., 1955). This main effect is modulated by one two-way interaction; for short vowels a left palato-velar consonant leads to a higher value for $F_1$ ($F(3, 2935) = 6.16$, $p < 0.001$). Since there are only 12 observations involved in this category, this interaction will not be interpreted further. A likelihood ratio test confirmed the necessity of including the two random terms (speaker: $\chi^2(1) = 259.6$, $p < 0.0001$, word: $\chi^2(1) = 207.7$, $p < 0.0001$). Overall this model suggests that shorter vowels do not differ from longer vowels in their first formant; schwa remains a mid-vowel in terms of $F_1$.

The value for $F_2$ is 1452 Hz on average (1484 Hz for short schwas and 1426 for long schwas). The statistical model for $F_2$ shows three main effects: duration ($\beta = 31.7$, $t = 4.7$, $p < 0.0001$), right ($F(3, 2937) = 22.9$, $p < 0.0001$) and left ($F(3, 2937) = 109.0$, $p < 0.0001$) consonantal context. There is no significant interaction in this model. As expected, $F_2$ is higher for shorter vowels for all the consonantal contexts considered. Fig. 2 illustrates the changes in the first two formants as a function of duration according to the following consonant; there is a global shift toward higher $F_2$ values for shorter schwas (pointed by the arrow). A likelihood ratio test confirms the necessity of including the two random terms (speaker: $\chi^2(1) = 313.6$, $p < 0.0001$, word: $\chi^2(1) = 180.0$, $p < 0.0001$).

In sum, this investigation of over 3000 acoustically defined schwa variants thus revealed a non-negligible proportion of very short schwas. Furthermore, shorter schwas are spectrally reduced. They are more coarticulated and thus take on the spectral characteristics of the surrounding consonants.

4. Study 2: Perceptual categorization of schwa and non-schwa variants

The first goal of this study is to test whether deciding on the presence or absence of schwa in a word token based on perceptual grounds is as categorical as it is usually assumed. Two lines of evidence lead us to suspect that perceptually ambiguous tokens do exist and are not marginal. Firstly, in our previous work on French schwa, we recurrently came across occurrences for which the categorization of schwa as being present or absent was difficult. Secondly, several studies in other languages and for different segments have reported discrepancies between transcribers when they have to decide whether the segment is absent or present. For instance, in Pitt et al. (2005), when one of the four trained transcribers considered a segment as present, this segment was only considered present by the other transcribers in 86% of cases. In Kuijpers and van Donselaar (1997) transcribers disagreed on the presence of Dutch schwa in 17% of the occurrences. Hence, it would be surprising if, unlike other segments, the presence versus absence of schwa in a given occurrence were always unanimously decided.

The second objective of this study is to examine the acoustic properties of these potentially ambiguous items and the variables influencing perceptual judgements. Obviously duration is an important cue to the perception of a vowel. However, many studies have also reported the influence of other variables on phoneme identification and perception. According to Dupoux, Kakehi, Hirose, Pallier, and Mehler (1999) phonotactic knowledge may induce the perception of an absent vowel in pseudowords. Hallé, Chéreau, and Segui (2000) showed that orthographic knowledge also shapes the perception of consonants. Finally Kemps, Ernestus, Schreuder, and Baayen (2004) document the influence of contextual variables on the perception of reduced words. While little is known about the variables influencing the perception of the presence versus absence of a vowel in tokens extracted from natural, connected speech, it is likely that similar variables are used to guide perceptual judgements, especially for those occurrences for which acoustical evidence is unclear.

![Fig. 2. $F_1$ and $F_2$ values of schwas according to following consonantal context (labial, dento-alveolar, palato-velar, uvular) and duration of the vowel: The arrow goes from long (> 50 ms) to short (< 50 ms) schwas.](image-url)
4.1. Method

A single judge did an initial screening of the 4294 tokens. Based on perceptual information only, he had to categorize the tokens as being ‘clearly with schwa’, ‘clearly without schwa’, or ‘ambiguous’ (i.e., the presence versus absence of schwa was not clear). According to this three-way categorization 92% of the tokens were classified as belonging to one of the two variants (25% without schwa, 67% with schwa). The remaining 8% of the tokens (330 tokens) were classified as ambiguous. A subset of items was then selected from these 4294 tokens for a multi-judge categorization test. These included an equal number of ‘clearly without schwa’ tokens (24), ‘clearly with schwa’ tokens (24), and ‘ambiguous’ tokens (24). These 72 items (given in Appendix A) were selected based on their intelligibility in isolation, the absence of background noise, the diversity of lexical types (a lexical item appeared only once in the experiment), and the position of the schwa (fifteen initial and nine medial schwas in each category). Twenty-two French native speakers from the University of Geneva took part in this categorization test. They were presented with the list of 72 items three times. The items were randomized inside each list. For each item, two orthographic transcripts, one with the schwa and one with an apostrophe replacing the schwa, appeared on a computer screen for 2000 ms. An auditory presentation of the item was played twice via headphones 750 ms after the written transcripts had disappeared. Participants had to press a button labelled “avec e” (with e) headphones 750 ms after the written transcripts had disappeared. An auditory presentation of the item was played twice via replacing the schwa, appeared on a computer screen for 2000 ms.

4.2. Results

4.2.1. Single judge classification

The classification of the stimuli by the first judge was as follows: the ambiguous group consisted of 240 tokens of words with an initial schwa and 90 tokens of words with a medial schwa. Interestingly, 136 tokens had no acoustic interval attributable to schwa according to our acoustic criteria (voicing and formant structure), 194 tokens had a schwa acoustic interval with a mean duration of 28 ms (sd = 13). We ran a generalized mixed effect model with speaker and word as random terms and whether the token received an ambiguous or non ambiguous categorization (present–absent responses combined) as a response. Results showed that the probability for a token to be judged ambiguous is higher if the schwa is in the first syllable (F(1, 4282) = 10.4, p < 0.01) of the word, and if one of the consonants surrounding the schwa is a voiceless obstruent versus a voiced obstruent or sonorant (right consonant: F(2, 4282) = 13.7, p < 0.0001; left consonant: F(2, 4282) = 5.8, p < 0.01). Phonotactic constraints (whether the cluster formed by the consonants surrounding the schwa exists in French or not) had no influence.

4.2.2. Multi-judge classification

For each item, we obtained 66 judgements (three repetitions per token for each of the 22 participants). In order to determine whether an item had received unanimous or heterogeneous judgements across the 66 trials (i.e., whether the schwa was perceptually ambiguous), we conducted a χ² test for each item, comparing the distribution of the participants’ responses to two theoretical distributions containing only unanimous responses; a distribution with 100% “with schwa” responses and a distribution with 100% “without schwa” responses. Results showed that 26 out of the 72 items did not differ from a theoretical distribution of only “with schwa” responses and could thus be considered as classified unanimously as variants with schwa. Twenty-five items did not differ from a theoretical distribution of only “without schwa” responses and were thus considered as being unanimously judged as variants without schwa. The remaining 21 items differed from both theoretical distributions and were thus considered as ‘ambiguous’ items in terms of the presence/absence of schwa. The classification of the items according to the participants’ responses in the categories “with schwa”, “without schwa” and “ambiguous” does not correspond exactly to the categories defined by the first judge. Three items classified as ‘clearly without schwa’ by the first judge are ambiguous according to the 22 judges’ estimations and six items initially classified as ambiguous are clearly variants without schwa (three items) or clearly variants with schwa (three items) according to the 22 judges’ estimations. These items are in bold in Appendix A, which provides for each item the categorization of the first judge, the percentage of “with schwa” responses and the derived classification based on the 22 judges’ responses. On the whole, the agreement between the first single judge’s classification of the 72 items and the classification by the 22 judges in the perception test is high and significant (Spearman rho = 0.91, S = 5584.4, p < 0.0001). Nonetheless, the few discrepancies between the classification of the single judge and of the multi-judges suggest that listeners may differ in their judgements. This is further confirmed by the multi-judge classification. For all items classified as ambiguous on the basis of the 22 judges’ estimations, the heterogeneity in the responses is due to a discrepancy in the

Fig. 3. Schwa durations (ms) as a function of the percentage of “with schwa” responses. Crosses, unfilled circles and filled squares are used to represent the items according to the perceptually defined categories ‘with schwa’, ‘without schwa’, and ‘ambiguous’, respectively.
responses across participants as well as within the three responses to the same token for a given participant. Depending on the items, between 10% and 60% of the participants did not give three identical responses to the three presentations of the same token (40% on average). This heterogeneity confirms that the presence of schwa in some tokens is unclear.

We then examined the duration of schwa (i.e., voiced interval with formant structure) for the 72 test items, in order to determine how it related to perceptual responses. In Fig. 3, the distribution of the items is plotted against the raw percentage of ‘with schwa’ responses and the acoustic duration of the schwa. Three categories of items are distinguished, based on the result of the 22 judges’ responses: ‘with schwa’ (circles), ‘without schwa’ (crosses) and ‘ambiguous’ (dark squares) items.

Unsurprisingly, a positive correlation is observed between the duration of the schwa interval and the percentage of “with schwa” responses ($r = 0.8, p < 0.0001$). Items with a schwa longer than 51 ms always belong to the category “with schwa” and, for most of the items of the category “without schwa”, no voiced interval with formant structure (i.e., interval of 0 ms duration) was identified. However, the presence of a voiced interval with formant structure and the duration of this interval are not enough to explain the perceptual categorization. On the one hand, some items were categorized as non-schwa variants despite the presence of a voiced interval with formant structure in the acoustic record. One of these items is illustrated in Fig. 4a: besoin ‘need’ has a schwa of 41 ms but received few “with schwa” responses (4.5%). On the other hand, there is no correlation between the percentage of “with schwa”

![Fig. 4.](image-url) (a)–(e) Spectrograms and waveforms of five different items (duration of the measured interval with voicing and formant structure (ms) and the rate of ‘with schwa’ judgements) in the corpus: (a) besoin ‘need’ (41 ms, 4.5%), (b) demi ‘half’ (51 ms, 11%), (c) prévenu ‘warned’ (30 ms, 32%), (d) provenance ‘origin’ (23 ms, 62%), (e) provenant ‘originating’ (0 ms, 17%) (note that this filtered item is produced in a phone interview).
responses and schwa duration when this duration is below 51 ms ($r=0.2$, $p>0.05$). All the ‘ambiguous’ items show values below this duration. This is illustrated in Fig. 4b–d: in demi ‘half’, prévenu ‘warned’, and provenance ‘origin’, schwa durations decrease (51, 30 and 23 ms, respectively) while the rate of “with schwa” responses increases (11%, 32%, 62%, respectively). In the last example illustrated in Fig. 4e, the item provenant ‘originating from’ has no clear voiced formant structure attributable to schwa, but this item received 17% of “with schwa” responses.

While duration has, as expected, a major influence on the participants’ responses, it is apparently not the only variable influencing responses. In order to determine which other variables contributed to the perceptual categorization of the tokens, we ran a generalized mixed effect model with the participants’ response as the dependent variable (“with schwa” versus “without schwa”) and speaker and word as random terms. The following predictors were entered sequentially in the model and retained if significant (at $p<0.05$): schwa duration (linear and quadratic term), number of repetitions, speech rate, lexical frequency as given for films by the LEXIQUE database (New, Pallier, Ferrand, & Matos, 2001), sonority of the right consonant, sonority of the left consonant, whether the consonants surrounding schwa form a legal French cluster, word duration (ms), schwa position in word and the estimated frequency of the non-schwa variant (as given by Racine, 2007). Speech rate was defined as the number of syllables in the utterance containing the given word divided by the duration of the utterance (in ms). Utterances consisted of the two words preceding the target word, the target word itself, plus the two words following. Word(s) preceding or following pauses (depending on whether the pause was before or after the target word) were not taken into account. The interactions of these predictors with schwa duration were also entered in the model and retained if significant. When two predictors (A and B) were correlated, they were orthogonalized in order to avoid collinearity. Orthogonalization was performed by running a linear model predicting variable A as a function of variable B. The residuals of this model were then used as values for the predictor in the mixed model instead of the raw values for variable A. The final model shows that the probability of obtaining a “with schwa” response increases as expected with schwa duration ($\beta=0.17$, $F(1, 4739)=276.2$, $p<0.0001$) and decreases with speech rate (orthogonalized with word duration, $\beta=-0.028$, $F(1, 4739)=41.8$, $p<0.0001$). Furthermore, the probability of obtaining a “with schwa” response is influenced by right ($F(2, 4739)=3.6$, $p<0.05$) and left consonant contexts ($F(2, 4739)=47.4$, $p<0.0001$). It is higher when the right or left consonant is a sonorant rather than a voiceless obstruent. The model also shows three two-way interactions involving schwa duration: schwa duration by right consonant sonority ($F(2, 4739)=9.5$, $p<0.0001$), schwa duration by left consonant sonority ($F(2, 4739)=16.1$, $p<0.0001$) and schwa duration by word duration ($F(1, 4739)=4.6$, $p<0.05$). The two interactions involving the segmental context are illustrated in Fig. 5a and b. Including a random term for judge and word significantly improves the model according to likelihood ratio tests (judges’ goodness of fit: $\chi^2(1)=14.9$, $p<0.001$, word: $\chi^2(1)=138.2$, $p<0.0001$). Furthermore, judges differ in their sensitivity to speech rate but not to the other variables ($\chi^2(2)=9.1$, $p<0.05$).

The interaction between context and duration is similar for the right (Fig. 5a) and left (Fig. 5b) consonantal contexts. It shows a difference in the slope between voiceless obstruent contexts on one hand and voiced obstruent and sonorant contexts on the other hand. When the consonant is a voiceless obstruent, the absence of an acoustic interval attributable to schwa never leads to a “with schwa” response. However, as soon as a small interval with voicing and a formant structure is present, the probability of a “with schwa” response rises drastically in this context. With voiced obstruents and sonorant consonants, a longer duration of schwa is needed to increase in a similar way the probability of a “with schwa” response. It is likely that in this context, the presence of voicing and formant structure can be less directly attributable to the presence of a vowel. Listeners probably often consider these cues to be part of the voiced obstruent/sonorant consonant. The interaction between schwa duration and word duration shows that for shorter words, a longer schwa interval is needed to lead to a “with schwa” response than for longer words. It is known that segments are longer in shorter words (Lehiste, 1972), see also Léon (1966) who mentions a correlation between schwa duration and word length). Listeners probably adapt their interpretation of the acoustic cues as a function of the surrounding segments’ durations. As a consequence, for a schwa to be considered present, the duration of its acoustic cues can be shorter for longer words.

In sum, the results for this second study suggest that a non-negligible proportion of the occurrences of schwa words in connected speech is ambiguous regarding the presence of schwa when presented out of context. Furthermore, while schwa

![Figure 5](https://example.com/figure5.png)

Fig. 5. (a) and (b) Interactions between schwa duration and right consonant sonority (a) and between schwa duration and left consonant sonority (b) as predicted by the statistical model for responses.
duration (defined here as an interval with voicing and formant structure) is an important variable in deciding whether schwa is present or absent, other variables are also relevant.

5. Discussion and conclusion

Schwa alternation in French has generally been described in phonological terms as a categorical process resulting in two discrete outputs: a schwa and a non-schwa variant. In the vast socio-phonetic literature studying rate of schwa occurrence or the factors governing its production, the distinction between schwa and non-schwa variants has almost always been presented as straightforward and unambiguous. This view of schwa contrasts with a large body of research on segment reduction, showing that segments may be reduced or even disappear in connected speech. Unlike that of other French vowels, the variability of schwa has not been studied. The two studies in this paper have attempted to fill this gap by describing the acoustic properties of schwa variants and by studying the ability of judges to categorize schwa tokens in schwa versus non-schwa variants.

In the first study, we investigated the phonetic variability of schwa in a large number of occurrences of schwa variants and found that schwa, like other segments in French, undergoes phonetic reduction. From a temporal point of view, a non-negligible number (14%) of tokens are made up of only a short interval of voicing with formant structure (30 ms). From a spectral point of view, short shaws are found to be more affected than longer shaws by surrounding consonants. This behaviour mirrors the spectral variation of other vowels in the same corpus (Gendrot & Adda-Decker, in press). In our second study, we investigated the perception of schwa word tokens extracted from connected speech. Two interesting findings emerged. Firstly, we observed inter and intra-judge disagreements on deciding whether a token was produced with a schwa or without. According to a single judge’s categorization, 8% of 4294 occurrences of alternating words in our corpus are ambiguous in this respect. We further showed that listeners do not rely only on the acoustical properties of the tokens to make their judgements. Other variables such as speech rate, word length and segmental context also have an influence on the judgements.

These acoustic and perceptual results have important theoretical and methodological implications. They point to a possible difficulty in establishing the origin or nature of some non-schwa variants. Given the temporal reduction of schwa we observed in our first study, we hypothesize that extreme phonetic reduction of shaws could lead to the complete disappearance of the acoustic cues associated with that vowel in running speech. Su (2003) showed that reduction led to the complete disappearance of vowels (other than schwa) in 4% of words in spontaneous French. In a larger corpus of interviews (245k words analyzed with an automatic alignment system), Adda-Decker, Boula de Mareuil, Adda, and Lamel (2005) found a deletion rate of 6% for vowels other than schwa. It is likely that extreme acoustic reduction affects shaws at similar rates. If schwa indeed undergoes extreme temporal reduction, then occurrences without schwa may have two origins. Either these forms can be the result of a categorical process, or, alternatively, they could be the result of extreme phonetic reduction.

The question also arises as to whether the whole conception of schwa alternation as a categorical process has to be reconsidered in light of these findings about reduction. In other words, are non-schwa forms simply reduced forms? Throughout this paper, we have assumed a categorical process, arising in the phonological or lexical component of the production system, to account for the alternation between schwa and non-schwa variants (for arguments in favour of a lexical account see Bürki, Ernestus, & Frauenfelder, 2010).

We have postulated that this categorical alternation has to be distinguished from the phonetic reduction process affecting all segments in running speech. However, since the nineties, a large body of research has demonstrated that processes traditionally assumed to result from categorical phonological rules (such as assimilation or sandhi phenomena) may also be explained by gradual phonetic processes (e.g., Browman & Goldstein, 1992; Holst & Nolan, 1995). Within the framework of Articulatory Phonology, Smorodinsky (1998) and Barnes and Kvititskaya (2002) have suggested that schwa disappearance in French is the endpoint of such a gradient phonetic process; in non-schwa variants, schwa would be obscured by extreme overlap with adjacent consonantal gestures. Two major arguments can be raised against this view. Firstly, schwa alternation in French is not a fast speech or casual speech phenomenon per se. Non-schwa variants also occur at slow rates and in careful pronunciations (see also Côté & Morrison, 2007). Secondly, non-schwa variants in French are often associated with voice assimilation between the two consonants that become adjacent. This assimilation operates word internally (e.g., jeton [ʒetɔ̃] ‘counter’ is very frequently produced as [ʃetɔ̃]) and across word boundaries (e.g., je suis pos. [ʒoʃuɪs pɔs] ‘I don’t know’ often realized [ʃoʃuɪs]). Such voicing requires that the laryngeal configuration at the transition between the consonants does not adjust to the configuration of an intervening schwa vowel. Explaining this assimilation by an overlap would require that the glottal gesture of the consonant following schwa hides both schwa and its preceding consonant. From our point of view these facts are better explained by a categorical alternation account rather than by a gradual overlap account.

The results of our two studies, documenting the existence of acoustically reduced and perceptually ambiguous tokens, also raise methodological questions related to the categorization of tokens as schwa versus non-schwa variants. Overall, our studies suggest that the methodological choices made for categorization may have important consequences on what is counted as an occurrence of one or the other variant. A first methodological option relates to the overall method: perceptually and/or acoustically based judgements. Depending on the method used, the categorization outcome for reduced tokens will probably differ. In our first study, we based ourselves on the acoustic properties of the tokens and considered tokens with a short voiced interval with a formant structure as schwa variants. However, if a categorization by ear had been applied to these data, these reduced tokens would have likely been categorized as non-schwa variants or as ambiguous. Almost all the tokens with a schwa interval shorter than 30 ms included in our perception study were indeed categorized as non-schwa variants or have received heterogeneous categorization.

In addition, each categorization method (perceptual and acoustic) comes with its own methodological options, which are also likely to influence the categorization outcome. If an acoustic method is applied, the acoustic criteria used for categorization will be crucial. For example, in our study, we based ourselves on the presence of a period of voicing with formant structure to define schwa variants. While this criterion is the least controversial cue for defining a vowel and is easy to implement for such a large amount of tokens, it has the drawback of excluding potentially devoiced shaws and of including potential excrescent transitional vocoids as schwa variants. One could argue that the proportion of schwa variants in our study (28%) may have been over- or under-estimated by the acoustic criteria we chose.

If a perceptual based categorization is performed, the categorization output will be influenced by the amount of contextual information available during the categorization process. We have shown that even when presented with the word in isolation, listeners rely on other criteria than schwa duration to make their perceptual judgement. Our analyses show that speech rate, word length and
segmental context have an influence. This result suggests that the speakers’ expectations play a role in their perception of the schwa. The influence of the speaker’s expectations in vowel perception for words presented in isolation is also observed by Connine, Ranbom, and Patterson (2008) for schwa words in English. In a syllable counting task, they observe that participants are influenced by the frequency of the variants when judging ambiguous schwa tokens. Similarly, Pitt (1998) showed that phonotactic constraints (legal versus illegal sequences of segments) influence the perceptibility of schwa in English words. In the same vein, Spinelli and Gros-Balthazard (2007) showed that phonotactic knowledge could help listeners to restore the schwa in a word recognition task.

As the amount of contextual information for a token increases, the effect of a listener’s expectations on the categorization will probably also increase. One could argue that the relatively high proportion of ambiguous cases we obtained in the perceptual test is linked to the restricted context in which the tokens are judged. In our perceptual study, the items to be judged were produced by different speakers and were presented in isolation. In many studies, perceptual judgements on schwa and non-schwa variants are made on long sequences of speech produced by the same speaker (see for example the work on the PFC corpus, Durand, Lèles, & Lyche, 2005). In these conditions, a coder will be more likely to expect a given variant if he/she is familiar with the speaker and if he/she can rely on contextual information. In the PFC corpus for instance, only 5 out of 5166 occurrences of word initial schwas were rated as uncertain (0.01%).

Several studies have shown that context may help to restore missing or less salient segmental information in perception tasks. For instance, Arai (1999) compared the perception of speech segments when embedded in long versus short sequences. He showed that the segment of speech is perceived completely differently in the two contexts. Similarly, Kemps et al. (2004) report that listeners can restore reduced forms in Dutch, but only when heard in the appropriate context. In our study, the amount of available contextual information is restricted, and the judges are probably more focused on the available acoustic information. When this information is ambiguous, for example, when the vowel is very short, or when the voiced interval could be interpreted as a part of the surrounding voiced consonant, then the judgement may be less clear cut.

Overall, these methodological considerations strongly suggest that the options taken to perform a categorization of the tokens are likely to have a strong influence on what is counted as a schwa or non-schwa variant. Our point here is not to propose a single best option. However, since the resulting categorization is likely to influence the studies outcomes, it seems crucial that authors explicitly report their methodological choice and discuss their possible consequences.

To conclude, our two studies have shown that the conception of French schwa alternation as a mere categorical process, leading to two distinct outputs (schwa and non-schwa variants) is too simplistic. Schwa in connected speech also undergoes phonetic reduction, leading to tokens with few or no acoustic cues signalling its presence. The situation becomes even more complicated if one considers that some of the very short schwas that surface in the acoustic outputs belong to a third category: excrescent schwa-like vocoids arising for articulatory reasons between some consonants (as in avr’l ‘april’ [avr’l]). How to differentiate between these types of surface forms (categorically deleted schwa, phonetically reduced schwa and excrescent vowels) and determine their origin is an important question that remains to be resolved. In order to do so, further studies should investigate in detail the acoustic properties and variability of non-schwa tokens which have not been analyzed in this study.

**Acknowledgements**

The authors would like to thank the editors and the three anonymous reviewers for their useful comments on previous versions of this paper, and Guillaume Gravier for the automatic alignment of the corpus.

**Appendix A**

See Table A1.

---

**Table A1**

List of the items used in the multi-judge classification test presented in study 2.

In column (1), classification of the items as clearly with schwa (+), clearly without schwa (−) or ambiguous (A) by the single judge; in (2), percent of “with-schwa” responses to the items over the 66 judgements (3 repetitions * 22 judges); in (3), classification of the items derived from the distribution of the responses in the multi-judge experiment as with schwa (+), without schwa (−) or ambiguous (A).

<table>
<thead>
<tr>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Besoin</td>
<td>A 5% A</td>
<td>100% +</td>
</tr>
<tr>
<td>Céla</td>
<td>A 89% +</td>
<td>100% +</td>
</tr>
<tr>
<td>Cheval</td>
<td>A 8% –</td>
<td>100% +</td>
</tr>
<tr>
<td>Deïla</td>
<td>A 0% –</td>
<td>100% +</td>
</tr>
<tr>
<td>Demandé</td>
<td>A 59% A</td>
<td>100% +</td>
</tr>
<tr>
<td>Dgmi</td>
<td>A 11% A</td>
<td>97% +</td>
</tr>
<tr>
<td>Recours</td>
<td>A 100% A</td>
<td>100% +</td>
</tr>
<tr>
<td>Regroupe</td>
<td>A 80% A</td>
<td>97% +</td>
</tr>
<tr>
<td>Relation</td>
<td>A 67% A</td>
<td>99% +</td>
</tr>
<tr>
<td>Remonté</td>
<td>A 92% A</td>
<td>99% +</td>
</tr>
<tr>
<td>Agéndra</td>
<td>A 50% A</td>
<td>100% +</td>
</tr>
<tr>
<td>Révue</td>
<td>A 30% A</td>
<td>100% +</td>
</tr>
<tr>
<td>Sélion</td>
<td>A 85% A</td>
<td>100% +</td>
</tr>
<tr>
<td>Séra</td>
<td>A 68% A</td>
<td>100% +</td>
</tr>
<tr>
<td>Vénir</td>
<td>A 0% –</td>
<td>100% +</td>
</tr>
<tr>
<td>Appelé</td>
<td>A 71% A</td>
<td>100% +</td>
</tr>
<tr>
<td>Biévue</td>
<td>A 65% A</td>
<td>99% +</td>
</tr>
<tr>
<td>Dangerex</td>
<td>A 18% A</td>
<td>100% +</td>
</tr>
<tr>
<td>Enlèves</td>
<td>A 32% A</td>
<td>100% +</td>
</tr>
<tr>
<td>Enégâtir</td>
<td>A 46% A</td>
<td>100% +</td>
</tr>
<tr>
<td>Prévenu</td>
<td>A 32% A</td>
<td>99% +</td>
</tr>
<tr>
<td>Provenance</td>
<td>A 62% A</td>
<td>99% +</td>
</tr>
<tr>
<td>Souvenir</td>
<td>A 14% A</td>
<td>99% +</td>
</tr>
<tr>
<td>Souvené</td>
<td>A 15% A</td>
<td>100% +</td>
</tr>
</tbody>
</table>
References
Gendrot, C., & Adda-Decker, M. Impact of duration and consonantal context on formant values of oral vowels: An automatic analysis of large broadcast news corpora in French. In M. Embarki, & C. Dodane (Eds.), Patterns of speech articulation: Between physics and metaphysics. Cambridge: CUP, in press.