

# Phoneme deletion in spontaneous British English

Olivier Corbin

Laboratoire de Phonétique et Phonologie (UMR 7018), CNRS / Sorbonne-Nouvelle, Paris, France

E-mail: oliviercorb@netcourrier.com

## ABSTRACT

The purpose of this acoustic study was to investigate the differences in phoneme deletion between lexical and function words in spontaneous British English and, where possible, compare our results with parallel studies conducted on French and Taiwanese Mandarin [1]. We also made an inventory of deleted consonants and vowels with respect to the syllable structure. Four native speakers of British English were recorded while having spontaneous conversations in order to obtain fifteen minutes of corpus. The global results show that consonants are more likely to disappear (5.3%) than vowels (2.4%) and that vowels are more often deleted in lexical words than in function words while it is the reverse for consonants. Alveolar consonants represent 76% of the deleted consonants and all deleted vowels belong to the non-low lax category.

## 1. INTRODUCTION

Phoneme deletion causes a particular problem for phonetic (phonemic) transcription of large databases, a necessary step for language comparison. This study is part of a larger program aiming to compare phoneme deletion in spontaneous speech in various languages (French, Taiwanese Mandarin, British English). The aim of this program is to contribute to a better cross-linguistic understanding of similarities and differences concerning phoneme deletion in spontaneous speech. In the published literature, few studies have been conducted on phoneme deletion in British English. The current study helps to fill this gap. Particular attention was given to the role of lexical versus function words, patterns of deletion in consonants versus vowels and the role of syllable structure as a determining context.

## 2. METHODOLOGY

The corpus consists of 15 minutes of spontaneous British English uttered by 4 native speakers (2 males and 2 females) in their mid-twenties whose accent can be regarded as near-RP. The two dialogues were recorded in a sound-proof room after a reading task, which was also recorded. The speakers were not aware at the time that the recording continued after this reading task. The acoustic signal was digitised and then spectrograms were produced and segmented manually using Praat software [2]. A very conservative criteria was used to determine deletion. Only those phonemes which did not correspond to a particular segment on the spectrogram and for which no locus nor obvious

evidence of assimilation were visible were considered deleted. Following Su & Basset [1], we investigated the percentage of deleted phonemes, the percentage of vowel deletion versus consonant deletion, the influence of the class of words (function or lexical words) and finally the influence of syllable structure.

## 3. RESULTS AND DISCUSSION

### 3.1 Global results

Table 1 shows the results of our study compared to the results obtained by Su & Basset (op. cit.). With 4.24% of phonemes deleted, it seems that phoneme deletion is slightly more important in British English than in French and Taiwanese Mandarin, at least in the studied corpora.

	Total	Deleted	%
<i>British English</i>	7637	324	4.24
<i>French</i>	10558	372	3.5
<i>Taiwanese Mandarin</i>	4780	175	3.5

Table 1: Total versus deleted phonemes

### 3.2 Lexical versus function words

It is widely believed that phoneme deletion affects more heavily function words than lexical words and familiar words than unfamiliar words [3].

In order to check this assumption, we separated the phonemes of our corpus in two categories depending on the category of the word they belong to (function or lexical). Nouns, adjectives, verbs and adverbs belong to the lexical category whereas pronouns, auxiliaries, articles, conjunctions and prepositions belong to the function words category. It turned out that approximately 58% of the words in our corpus were function words. Without further study, it is not possible to determine whether the similarity between percentage of function words in the corpus and percentage of deletion in function words is meaningful.

Table 2 shows that a majority of deleted phonemes come from function words (60.5% in function words versus 39.5% in lexical words). Table 3 shows that phoneme deletion is more likely to happen in function words than in lexical words (5.9% versus 2.9%).

	Total	%
Deleted	324	100
Function	196	60.5
Lexical	128	39.5

**Table 2:** Deleted phonemes in lexical and function words.

Those results are noticeably different to those obtained on French and Taiwanese Mandarin, where Su & Basset obtained 75% of deletion occurred in function words. This may be due to the fact that English has a lexical stress which results in an important reduction of the unstressed syllables.

	Total	Deleted	%
Function	3272	196	5.9
Lexical	4365	128	2.9

**Table 3:** Deleted phonemes in function words versus lexical words

Here again, if the percentage of deletion in function words is comparable with those obtained for French and Taiwanese (respectively 8% and 5.5%), the percentage for lexical words (2.9%) is considerably higher (respectively 1.3% and 1.7%).

### 3.3 Vowels versus consonants

Table 4 shows the results of consonant and vowel deletion in English, French and Taiwanese Mandarin. It seems as though -at least in our corpus- that consonants are more likely to be deleted than vowels in British English (5.3% of consonants versus 2.4% of vowels). It seems to be the same tendency in Taiwanese Mandarin - where the vowel bears the tone- but the reverse in French.

	Deleted (Total)		
	Brit. Eng.	French	Tai. Man.
Deleted C	5.3% (4709)	1.9% (5288)	4.4% (1766)
Deleted V	2.4% (2928)	5.3% (4933)	2.9% (2307)

**Table 4:** Percentages of deleted consonants and vowels in British English, French and Taiwanese Mandarin

We observe in table 5 that a great majority of consonants are deleted in function words while a majority of vowels are deleted in lexical words. This may be due to the fact that a great number of function words are monosyllabic, hence the tendency seems to drop the final consonant (very often) or the initial consonant (rather rare).

	Consonants	Vowels
Function	172	24
Lexical	80	48

**Table 5:** Number of deleted vowels and consonants in relation with the word category

#### 3.3.1 Vowels

In table 6 we show the vowels concerned by deletion. Three salient facts appear:

- (i) Unsurprisingly, [ə] is the most deleted vowel in our corpus (13%),
- (ii) Only five different vowels are concerned by deletion: [ɪ, ʊ, ə, ʌ, ɛ], all of them being non-low lax vowels,
- (iii) No low vowels, no tense vowels and no diphthongs have been totally deleted in our corpus.

Deleted/Total (%)

[ɪ] 17/599 (2.8)	[ə] 39/296 (13.2)	[ʊ] 3/63 (4.7)
[ɛ] 7/241 (2.9)	[ʌ] 4/161 (2.5)	

**Table 6:** Deleted versus total number of vowels

The fact that the central vowel is likely to be deleted seems to be shared by the three languages taken into account. But in contrast to the role of height in characterising the tendency of which vowels are likely to delete is common to British English and Taiwanese Mandarin, all vowels can undergo deletion in French.

#### 3.3.2 Consonants

Table 7 illustrates the results of deleted versus the total number of consonants. Some conclusions may be drawn from those statistics:

- (i) The deletion phenomenon concerns a wide range of consonants (13 different consonants)
- (ii) The most deleted consonant is [h], with 25% deletion, but notably 75% of them occurred in function words like *his*, *him*, *has* or *have* which did not bear any accent.
- (iii) A very large percentage of alveolar consonants are deleted while very few labial and velar consonants, as expected from Kohler [4]
- (iv) Fricatives, nasals and affricates seem to be quite resistant to deletion, though the two interdental [ð, θ] may be deleted in certain circumstances. All the [ð] occurred in function words like *them* or *that* and all

[θ] occurred in lexical words such as *something*.

Deleted/Total (%)

[p] 4/103 (3.8)	[t] 83/533 (15.5)	
	[d] 40/343 (11.6)	[g] 4/80 (5)
	[s] 13/363 (3.5)	
[v] 5/146 (3.4)	[z] 10/277 (3.6)	
	[θ] 3/59 (5)	
	[ð] 7/137 (5)	
[m] 9/278 (3.2)	[n] 17/487 (3.4)	
	[l] 29/228 (12.7)	[h] 30/116 (25.8)

**Table 7:** Deleted versus total number of consonants

It must be added that the frequency of words has a strong effect on deletion. For instance, the deletion of [t] is very often due to the loss of the final part of the negation on an auxiliary like in *doesn't*, *don't*, *can't* and sometimes due to very common words like *just*, *but*, *about*. The deletion of [d] is often due to the loss of the past morpheme of a verb but, in function words, the word *and* is responsible for all the deletions. Four forms coexist in the corpus: [ænd, æn, ən, ŋ], this may confirm the hypothesis of multiple representations of common words in the mental lexicon [5].

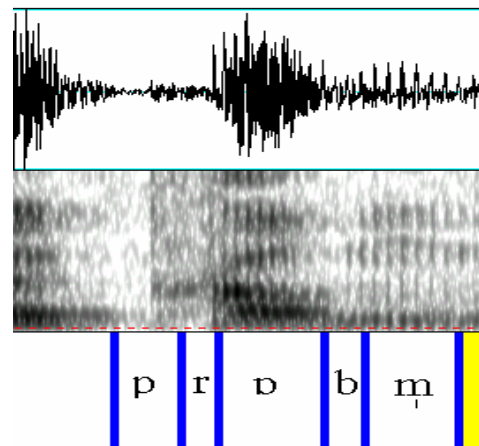
## 4. SYLLABLE POSITION AND DELETION

### 4.1 Deletion in onset of syllable

Among the 252 deleted consonants in our corpus, only 52 (23%) were deleted in attack position of the syllable. Only four different phonemes ([h, ð, θ, l]) may be deleted from this position and 80% of those deletions occurred in function words. The three [θ] deletions occurred in the word *something* which was realised as [sʌmɪŋ]. The deletions of [ð] only occurred in function words with the loss of the initial consonant in *them*, *that*. Finally, the deletion of [h] occurred mainly in 3<sup>rd</sup> person pronouns *he*, *him*, *her*, *his* and sometimes in the auxiliary *have*.

The situation is slightly different for [l], which can be deleted in both onset and coda positions. We have examples such as: *feel comfortable* which is realised as [fi:kʌmfɪtəb], showing that [l] can be deleted in the coda, but also the word *problem* which is realised several times as [prɒbəm] or even [prɒbm̩], showing that it can be deleted when in a complex onset of a

syllable (see Figure 1 below). We will return to the cluster simplification problem in 4.2.



**Figure 1:** Spectrogram of the word *problem* realised [prɒbm̩] by speaker 2

### 4.2 Deletion in coda of syllable

The vast majority of consonant deletions occurred from coda position (77%). We can distinguish three different types of deletion. The first is the simple deletion of the final post-vocalic consonant. It occurs very often in function words (*about* can be realised [ə'bau], for instance) but also in lexical words (*forget* can be realised [fə'ge]). The second is cluster simplification, where the last consonant disappears in a homorganic pair. This simplification affects almost only [t] and [d]. In function words, many auxiliaries used in a negative form are realised without the final [t], like [dəʊzn] for *doesn't* and in lexical words, many verbs in the preterit are realised without the final [d], like [faʊn] for *found*. This type of deletion is probably the result of a total assimilation of the stop followed by a shortening of the resulting long consonant. This process has been described by Nolan [6]. This is probably also the case when a verb which ending is a homorganic cluster receives the 3<sup>rd</sup> person inflection morpheme. In that case, the morpheme is never deleted but the final consonant of the root is very often absent, as in *sounds* which is realised [saʊnz]. The last type concerns the consonant [l], which seems to have a more variable behaviour. A word like *child* can be reduced in two different manners: [tʃaɪd] or [tʃaɪl], and we also found in our corpus two different reduced forms for the word *difficult* : [dɪfɪkʌt] and [dɪfɪkʌl]. Unfortunately, the small number of deletion of this type prevents us from drawing any conclusion on this variable behaviour.

### 4.3 Vowel deletion

As we mentioned above, only five different vowels have been found to delete in our corpus. No vowel was

deleted in a stressed syllable. In a great majority of cases, the deleted vowel belonged to an open syllable, the deletion giving birth to a consonant cluster as in [junrv3:stɪ] (which was found for the realisation of the word *university* on four occasions by two different speakers) or as in [spəʊz] (realisation of the word *suppose*). As we can see, these consonant clusters do not violate allowable phonotactics of English but we also noticed that [ə] deletion can give birth to forms such as [kmpɫɪtɪ] for the word *completely*. In that case, it appears that the [m] becomes syllabic. The same situation appeared for the word *problem* (see Fig. 1). Here, there might be an interaction between the [m] deletion the [l] deletion in order to avoid an ill-formed syllable. We can conclude that the conditions for vowel deletion are threefold:

- i) The vowel is unstressed,
- ii) The vowel is a non-low lax vowel,
- iii) The deletion does not entail an ill-formed syllable.

## 5. CONCLUSION

The results seem to show that (i) Phoneme deletion in spontaneous British English is slightly more important than in French and Mandarin Chinese, (ii) A majority of phonemes are deleted in function words, (iii) Consonants are more affected by deletion than vowels in function words, but it is the reverse for lexical words (iv) Alveolar consonants and [h] are more likely to be deleted than other consonants, (v) Only unstressed non-low lax vowels can be deleted and (vi) If we take the position where a consonant is more likely to disappear as weak, both coda and post-stress consonants are weak.

Moreover, this study has shown that phoneme deletion in spontaneous British English is somewhat different from phoneme deletion in French and Taiwanese Mandarin. This is partly due to the differences in syllable structure but also to the nature of the three compared languages. Nevertheless, it seems as though the tendency to have central/close vowel deletion is shared by the three languages. We are aware of the fact that “lenitions are less complete than they first appear” [7], and our results will have to be confirmed by perceptive experiments in a near future.

## ACKNOWLEDGEMENTS

I would like to thank Abigail Cohn and Angélique Amelot for their useful comments and my colleagues Patricia Basset and Tzu-Ting Su for sharing their data.

## REFERENCES

- [1] T. SU and P. Basset, “Language dependent and independent spontaneous speech phenomena”. Actes du séminaire international, SPOSS, Sound Patterns of Spontaneous Speech: Production and Perception, D. Duez, (ed.), Aix-en-Provence, 24 - 26 Septembre 1998. Aix-en-Provence, France.
- [2] Boersma, P. and Weenink, D. Praat, a system for doing phonetics by computer. Institute of Phonetic Sciences of the University of Amsterdam, report 132-182.
- [3] C. Bally, *Le langage et la vie*, Payot, Paris, p.148, quoted by Hagège, C. and Haudricourt, A. *La phonologie panchronique*, PUF, Paris, p.115, 1978
- [4] K.J. Kohler, “Segmental reduction in connected speech in German: Phonological facts and phonetic explanations” in *Speech Production and Speech Modelling*, W.J.Hardcastle & A. Marchal Eds., pp. 69-92. Dordrecht: Kluwer, 1990
- [5] J.J. Ohala, “The Relation Between Sound Change and Connected Speech Processes”. in *Sound Patterns of Connected Speech: Description, Models, and Explanations*, A. P. Simpson & M Pätzold Eds., pp.201-207 *Arbeitsberichte des Instituts für Phonetik und digitale Sprachverarbeitung der Universität Kiel (AIPUK)* 31, 1996.
- [6] F. Nolan, *Overview of English Connected Speech Process (CSPs)*, AIPUK, vol 31, Kiel, p. 15-26, 1996
- [7] S.Y. Manuel, “Speakers nasalise [ð] after [n] but listeners still hear [ð]”, *Journal of Phonetics*, vol 23, pp.453-476, 1995