

Probabilistic effects on French [t] duration

Francisco Torreira & Mirjam Ernestus

Radboud Universiteit Nijmegen, The Netherlands
Max Planck Institute for Psycholinguistics, The Netherlands

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

Abstract

The present study shows that [t] consonants are affected by probabilistic factors in a syllable-timed language as French, and in spontaneous as well as in journalistic speech. Study 1 showed a word bigram frequency effect in spontaneous French, but its exact nature depended on the corpus on which the probabilistic measures were based. Study 2 investigated journalistic speech and showed an effect of the joint frequency of the test word and its following word. We discuss the possibility that these probabilistic effects are due to the speaker's planning of upcoming words, and to the speaker's adaptation to the listener's needs.

Index Terms: lexical probability, prosody, French, corpus research, acoustic reduction, speech rate.

1. Introduction

A number of recent corpus studies have shown that probabilistic factors such as the frequency of occurrence and contextual predictability of specific linguistic units (e.g. words, syllables, segments) may affect articulation [1, 2, 3, 4]. For instance, [1] found a shortening effect of the frequency of a word on the duration of several affixes and affix-internal segments drawn from a corpus of spontaneous Dutch. Similarly, [3] reports that English content and function words tend to be shorter the higher their lexical frequency and conditional probability (i.e. probability of a word given its preceding or following words). These studies suggest that more frequent and predictable linguistic materials tend to receive shorter and weaker pronunciations, a claim also known as the Probabilistic Reduction Hypothesis [2]. The present work extends this line of research by investigating if probabilistic effects can be found in French on the duration of intervocalic [t] consonants in two different speech styles.

Predictability effects on articulation have been claimed to arise during the planning process in speech production [3]. When speech planning is more difficult (e.g. because the words being planned are less probable), speakers may locally slow down their rate of articulation in order to preserve an undisturbed flow of speech. This hypothesis is supported by findings that predictability effects on content words are caused by the right, not the left, context (i.e. the following rather than the preceding word) [1, 3].

If word frequency and predictability effects arise during speech planning, it is expected that probabilistic effects will not be restricted to languages for which they have been established so far. Furthermore, it is expected that probabilistic effects will be greater in unprepared spontaneous speech than in read and prepared speech. In order to investigate these issues, we examined a new language and analyzed materials extracted from corpora of casual and journalistic speech.

The reason for investigating French is that the studies cited above, along with nearly all others, have exclusively focused on

Germanic languages. Importantly, Germanic and Romance languages differ in terms of their temporal structure. For instance, French and Spanish, which are said to be syllable-timed, exhibit less variability in syllabic durations within an utterance than Dutch, English and German [5]. As a consequence, probabilistic effects on the duration of syllables and segments might be smaller in French than in Germanic languages.

In Study 1 we extracted data from the Nijmegen Corpus of Casual French (NCCFr from now on), which contains highly casual conversations among Parisian university students [6]. In Study 2, we used speech materials extracted from the ESTER corpus of journalistic speech [7], which mostly contains read news broadcasts from several French-speaking radio stations.

We selected [t] as our object of study, since it can be found in numerous different French words representing a wide range of frequencies. Moreover, [t] closure, the main acoustic correlate of the duration of French [t], can be accurately measured from a minimally clean acoustic signal. In the light of the findings mentioned above, [t] closures can be expected to display shorter durations when occurring in more frequent or predictable words.

2. Study 1

In Study 1, we investigate whether probabilistic effects on articulation can be found in casual French.

2.1. Method

Excerpts containing intervocalic [t] consonants were extracted from the NCCFr [6], which includes 35 hours of conversation among friends. All [t] consonants occurred within words of minimally two syllables, and were preceded and followed by one of the vowels /a/, /e/, /ɛ/, /o/, /ɔ/. Word-medial positions were favored over word-initial and word-final positions in order to avoid word-edge effects. Words with [t] surrounded by high vowels were excluded in order to avoid cases of palatalization, which might affect the realization of closures. Nasal vowels were disconsidered in order to avoid cases of nasality spreading into [t] closures, a situation that would make our measurements of oral closures less accurate. Only word types with at least 5 tokens were considered for analysis. For word types that complied with all these conditions, we randomly extracted a maximum of 15 tokens. The average number of tokens per word type was 8.24, with a standard deviation of 4.03.

All word types were content words. We decided to discard two semantically weak words (*était*, and *toilement*) on the grounds that they might behave differently from the vast majority of word types [3]. Therefore, our materials exclusively consist of a homogeneous group of semantically strong content words. Examples of used word types include *arrêté*

/avɛtɛ/ ‘stopped’, *gâteaux* /gato/ ‘cake’, *hôtel* /otɛl/ ‘hotel’ and *photo* /foto/ ‘photograph’. The few tokens that displayed a partial closure were discarded, and only [t] consonants displaying full closures were retained for analysis. In total, our dataset contained 29 distinct word types and 239 tokens spoken by 44 different speakers.

The duration of [t] closures was measured manually from speech waveforms. We provided a prosodic annotation to each of the extracted tokens including pitch accents, intonational boundaries and the number of syllables of the word containing [t]. Pitch accents were annotated whenever they occurred on the vowels preceding and following [t]. No distinctions were made between different sorts of pitch accents. However, it may be noted that all accents on vowels preceding [t] were initial accents found at the beginning of accentual groups (also called *accents d’intensité* in the French intonational literature) [8], while accents on vowels following [t] could be either group initial or group final. This distribution follows from the fact that all of our [t] consonants were word-medial, and that final accents must always be word-final. Intonational boundaries were marked whenever a boundary tone occurred between the test word and preceding or following words. No distinctions were made between different levels of prosodic phrasing. The vast majority of test words were not preceded by an intonational boundary (97%), but most of them were followed by one (74%). The number of syllables in the test word was calculated automatically from canonical pronunciations. Finally, speech rate was computed as the number of syllables per second within a stretch of speech containing the test word free of pauses or disfluencies.

Our probabilistic measures consisted of word and word bigram frequencies, both normalized per million words (for word i : frequencies of w_i , $C(w_{i-1}w_i)$ and $C(w_iw_{i+1})$). Two sets of frequency measures were computed. The first set was computed using the transcriptions of the NCCFr corpus. Since this corpus may be judged rather small for the purpose of estimating word and word bigram frequencies (it only has around 550,000 word tokens), we also used probabilistic estimates computed jointly from the ESTER corpus and the PAROLE French Corpus [9], which both consist mostly of journalistic texts. These two journalistic corpora together contain over 15 million words. The first set of probabilistic measures is more appropriate for modeling casual speech (as it is based on casual speech data), while the second set may be more reliable given the bigger size of the corpus used for its estimation.

For the computation of word bigram frequencies, the different types of punctuation marks used in the transcriptions of the corpora were considered word types of their own. Word bigrams can therefore be composed of a test word and a punctuation mark. Whenever a test word was preceded or followed by a pause not marked by a punctuation mark, the corresponding word bigram frequency was left blank in our data matrix.

Finally, a check was performed in order to assess the reliability of our manual measurements and annotations. Eighty-nine tokens (15% of 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 assistant (97.2% of [t] closure duration measurements differed by 10 ms or less between the two annotators. Cohen’s κ values of .94 and .67 were respectively obtained for the annotation of a preceding and a following intonational boundary).

2.2. Results

Mixed-effects linear regression with speaker and word type as random factors was used in order to estimate the effects of the different predictors on [t] closure duration. All probabilistic variables were logarithmically transformed, since their distributions had long right tails spreading over several orders of magnitude, and since language users are known to be sensitive to logarithmic values rather than raw frequencies. For ease of exposition, we do not continuously refer to these variables as log-transformed variables and rather use their original names (e.g. word frequency instead of log word frequency). We first fitted a control model with speech rate and the prosodic factors present in our annotation, and then added separately each of the two sets of probabilistic estimates. Whenever two predictors, say A and B, were significantly correlated, we orthogonalized them by replacing variable A in the regression model with the residuals of a linear model in which A was predicted by B. We only report the results of models which have been trimmed of outliers (i.e. of data points with values beyond two residual standard errors of the values predicted by the model).

2.2.1. Control model

Unsurprisingly, speech rate was a statistically significant predictor of [t] closure duration ($\beta = -4.25, t = -5.92, p < .0001$). From the set of prosodic factors considered in the annotation, only the presence of an accent on the vowel preceding [t] and the number of syllables in the word yielded statistically significant effects (preceding accent: $\beta = 8.63, t = 2.56, p < .05$; number of syllables.: $\beta = -5.3, t = -3.04, p < .005$). All other factors yielded regression coefficients in the expected direction, but none of them approached statistical significance. Only speech rate, the presence of a preceding accent and the number of syllables in the test word were retained in the model, to which we then added the probabilistic predictors.

2.2.2. Probabilistic factors

From the set of probabilistic factors estimated from the NCCFr, only word bigram frequency $C(w_iw_{i+1})$, that is the joint frequency of the test word and its following word, approached statistical significance ($\beta = -2.06, t = -1.81, p = .07$): there was a tendency for [t] closures to be shorter when occurring in more frequent word bigrams. Word frequency and word bigram frequency $C(w_{i-1}w_i)$ did not appear to affect [t] closure duration (word frequency: $\beta = -0.96, t = -0.6, p = .54$; word bigram frequency $C(w_iw_{i+1})$: $\beta = -0.5, t = -0.54, p = .58$).

From the set of probabilistic factors estimated from the journalistic corpora, only word bigram frequency $C(w_{i-1}w_i)$, that is the joint frequency of the test word and its *preceding* word, reached statistical significance ($\beta = -1.19, t = -2.13, p < .05$): the duration of [t] closures tended to be shorter the more frequent the word bigram. This tendency is illustrated in Figure 1. Word frequency and word bigram frequency $C(w_iw_{i+1})$ did not yield statistically significant effects on [t] closure duration (word frequency: $\beta = -1.07, t = -1.63, p = .1$; word bigram frequency $C(w_iw_{i+1})$: $\beta = -0.32, t = -0.4, p = .68$).

2.3. Discussion

In Study 1, effects of word bigram frequency on the duration of [t] closures were identified. These results show that a syllable-timed language like French may also exhibit probabilistic effects on articulation. Interestingly, different results were ob-

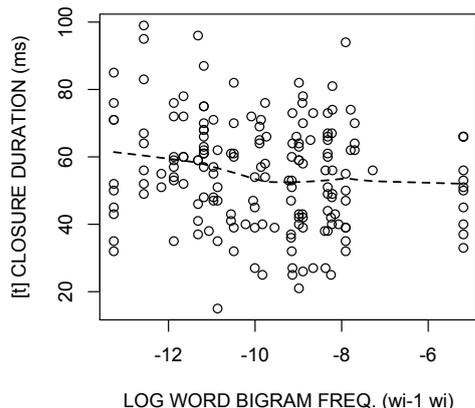


Figure 1: *Study 1: duration of [t] closure as a function of log word bigram frequency $C(w_{i-1}w_i)$ as estimated from journalistic sources. The dashed line represents a lowess regression curve.*

tained depending on which probabilistic estimates were used. An effect of word bigram frequency $C(w_{i-1}w_i)$ was identified when the probabilistic estimates were computed from a large corpus of journalistic texts. On the other hand, when we used probabilistic estimates computed from the NCCFr, a marginal effect of word bigram frequency $C(w_iw_{i+1})$ was found. It may be argued that the journalistic speech estimates used in Study 1 are not appropriate for modeling casual pronunciations and that the effect of word bigram frequency $C(w_{i-1}w_i)$ is not genuine, but rather arises from an underlying correlation between word bigram frequency $C(w_{i-1}w_i)$ and word bigram frequency $C(w_iw_{i+1})$. The reason why the latter did not yield a statistical effect might be attributed to the fact that it was poorly estimated. Another possibility is that both word bigram effects are genuine.

3. Study 2

In Study 2, we investigated whether probabilistic effects could also be found in journalistic speech, where speech planning may be less difficult and affect articulation less.

3.1. Method

The methodology used in Study 2 was similar to that used in Study 1, except for that the materials were extracted from the ESTER corpus of journalistic speech. Possible word types complied with the criteria exposed in Section 2.1. Consequently, the dataset consisted of semantically strong content words with [t] consonants in word-medial position surrounded by non-high oral unrounded vowels. As in Study 1, only word types with at least 5 tokens were considered for analysis, and a maximum of 15 tokens was used for each word type. Examples of used word types include *atomique* /atɔ̃mik/ ‘atomic’, *attaque* /atak/ ‘attack’, *hôtel* /otɛl/ ‘hotel’ and *bateau* /bato/ ‘ship’. The resulting dataset consisted of 63 word types and 592 tokens spoken by 376 speakers.

The annotation procedure was identical to that used in Study 1. The probabilistic measures were estimated only from

the journalistic PAROLE and ESTER corpora, which were deemed to be appropriate in terms of register and size. We decided not to use frequency measures computed from the NCCFr, since this corpus was relatively small and inappropriate in terms of register. As in Study 1, the phrasal position of the test words was not balanced. Most of the word tokens (72%) were followed by an intonational boundary, while only a few of them were preceded by one (6%). Again, we performed a reliability check on 15% of the tokens with the help of a second annotator. Inter-annotator agreement was satisfactory (95.4% of [t] closure duration measurements differed by 10 ms or less between the two annotators. Cohen’s κ values of .88 and .57 were respectively obtained for the annotation of a preceding and a following intonational boundary).

3.2. Results

As in Study 1, mixed-effects linear regression with speaker and word type as random factors was employed to identify statistically significant predictors of [t] closure duration and estimate their effects. Again, we first fitted a control model with speech rate and prosodic factors, and then added the probabilistic measures to this model.

3.2.1. Control model

Table 1 summarizes the regression coefficients, t values and p values of the statistically significant predictors. All control factors, except the presence of a preceding intonational boundary, yielded statistically significant effects in the expected direction. Speech rate and the number of syllables in the word were negatively correlated with closure duration. On the other hand, the presence of neighboring accents and of a following intonational boundary tended to increase closure duration.

Table 1: *Control model in Study 2: Regression coefficients along with t and p values.*

Factor	$\hat{\beta}$	t	p
Speech rate	-4.58	-7.77	< .0001
Fol. accent	12.67	6.57	< .0001
Fol. inton. bound.	9.31	5.32	< .0001
Prec. accent	4.89	2.23	< .05
Number of syl.	-4.54	-3.25	< .005

3.2.2. Probabilistic factors

Word bigram frequency $C(w_iw_{i+1})$ yielded a statistically significant effect ($\beta = -1.12, t = -2.57, p < .05$). Along the lines of the Probabilistic Reduction Hypothesis, [t] closures tended to be shorter the more frequent the word bigram (see Figure 2). On the other hand, word frequency and word bigram frequency $C(w_{i-1}w_i)$ did not have a statistically significant effect on [t] closure duration (word frequency: $\beta = -0.84, t = -0.81, p = .41$; word bigram frequency $C(w_{i-1}w_i)$: $\beta = -0.27, t = -0.63, p = .52$).

Since word bigram frequency can be expected to be lower across stronger prosodic boundaries, there is a possibility that the effect of word bigram frequency $C(w_iw_{i+1})$ is in fact a prosodic effect, since in our annotation we only coded one level of intonational phrasing. In order to address this issue, we tested the effect of word bigram frequency $C(w_iw_{i+1})$ on the subset of test words not followed by an intonational boundary.

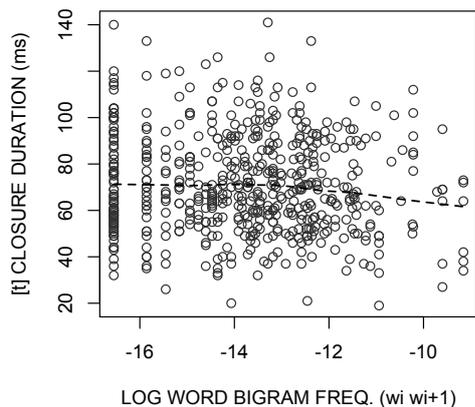


Figure 2: Study 2: duration of [t] closure as a function of log word bigram frequency $C(w_i w_{i+1})$. The dashed line represents a lowess regression curve.

The effect approached statistical significance ($\beta = -1.04$, $t = -1.81$, $p = .07$). The fact that the regression coefficient obtained for word bigram frequency $C(w_i w_{i+1})$ in this model remained similar to that obtained using the full dataset suggests that the loss of statistical significance may be due to the reduced sample size ($n = 140$; 23% of the full dataset).

4. General discussion and conclusions

This study investigated whether a word's predictability affects the duration of its [t] consonants in connected French. In Study 1, which used casual speech materials, an effect of word bigram frequency $C(w_{i-1} w_i)$ was identified when the probabilistic estimates were computed from a large corpus of journalistic texts. When we used probabilistic estimates computed from casual speech materials, a marginal effect of word bigram frequency $C(w_i w_{i+1})$ was found. In Study 2, which used journalistic speech materials and probabilistic estimates, we found an effect of word bigram frequency $C(w_i w_{i+1})$. In all cases, the duration of [t] closures tended to be shorter the higher the word bigram frequency. We conclude that French, a syllable-timed language, is also subject to the sort of probabilistic effects on articulation which have been identified for Germanic languages [1, 2, 3], which are said to be stress-timed.

Study 1 shows the importance of using appropriate corpora to estimate the probabilities of linguistic materials. This poses a serious challenge to researchers investigating casual speech, since for most languages no sufficiently large corpora of casual speech are available from which robust frequency estimates can be computed. Using inappropriate probabilistic measures may lead to erroneous or conflicting results.

Even though in Study 1 different results were obtained depending on which probabilistic measures were used (i.e. based on casual or journalistic speech), a consistent picture emerges if we only consider the results obtained with frequency estimates computed from corpora matched to the data in terms of speech style. In this case, we find that the duration of [t] closures in Studies 1 and 2 is sensitive to word bigram frequency $C(w_i w_{i+1})$, replicating previous findings that predictability ef-

fects for content words are essentially caused by the upcoming context [1, 3].

This interpretation supports the view that probabilistic effects are mainly due to speech planning constraints. To this respect, the results of Study 2 need some comment. Since our journalistic speech data mostly consisted of read texts, one may wonder to what extent the word bigram frequency effect found in Study 2 may be due to planning constraints. A possible explanation is that, even if planning constraints play a lesser role in read than in casual speech, the effect is strong enough to manifest itself in the speech of radio news readers.

Another possibility is that the probabilistic effect in Study 2 is not solely due to planning constraints, but also to adaptation on the part of news readers to their listeners' needs. Less frequent word sequences tend to be more informative from the point of view of the listener, and may therefore receive longer articulations. Although on-line adjustments of this sort are considered to be computationally complex and to take place slowly and imperfectly in connected speech [10], it is not inconceivable that professional news readers are able to successfully perform such adjustments. More research is needed to identify the sources of probabilistic effects on articulation in different speech registers.

5. Acknowledgements

We wish to thank Lou Boves, Kevin Russell and Christophe Dos Santos for their comments and help. This work was funded by a European Young Investigator Award to the second author.

6. References

- [1] M. Pluymaekers, M. Ernestus, and H. Baayen, "Articulatory planning is continuous and sensitive to informational redundancy," *Phonetica*, vol. 62, pp. 146–159, 2005.
- [2] A. Bell, D. Jurafsky, E. Fosler-Lussier, C. Girand, M. Gregory, and D. Gildea, "Effects of disfluencies, predictability, and utterance position on word form variation in English conversation," *Journal of the Acoustical Society of America*, vol. 113, no. 2, pp. 1001–1024, 2003.
- [3] A. Bell, J. Brenier, M. Gregory, C. Girand, and D. Jurafsky, "Predictability effects on durations of content and function words in conversational english," *Journal of Memory and Language*, pp. 92–111, 2009.
- [4] M. Aylett and A. Turk, "The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech," *Language and Speech*, vol. 47, no. 1, pp. 31–56, 2004.
- [5] E. Grabe and E. L. Low, "Durational variability in speech and the Rhythm Class Hypothesis," in *Laboratory Phonology Vol. 7*, C. Gussenhoven and N. Warner, Eds. Mouton De Gruyter, 2002, pp. 515–546.
- [6] F. Torreira, M. Adda-Decker, and M. Ernestus, "The Nijmegen Corpus of Casual French," submitted.
- [7] S. Galliano, E. Geoffrois, D. Mostefa, K. Choukri, J.-F. Bonastre, and J. Gravier, "ESTER Phase II Evaluation Campaign for the Rich Transcription of French Broadcast Newshase II Evaluation Campaign for the Rich Transcription of French Broadcast News," *Proc. Interspeech 2005*, pp. 2453–2456, 2005.
- [8] S.-A. Jun and C. Fougeron, "Realizations of accentual phrase in French," *Probus*, vol. 14, pp. 147–172, 2002.
- [9] PAROLE French Corpus. [Online]. Available: <http://catalog.elra.info/>
- [10] E. G. Bard, A. H. Anderson, C. Sotillo, M. Aylett, G. Doherty-Sneddon, and A. Newlands, "Controlling the intelligibility of referring expressions in dialogue," *Journal of Memory and Language*, vol. 42, pp. 1–22, 2000.