Disentangling perceptual and psycholinguistic factors in syntactic processing: Tone monitoring via ERPs

David J. Lobina¹ · Josep Demestre² · José E. García-Albea²

Published online: 13 July 2017
© Psychonomic Society, Inc. 2017

Abstract Franco, Gaillard, Cleeremans, and Destrebecqz (*Behavior Research Methods*, 47, 1393–1403, 2015), in a study on statistical learning employing the click-detection paradigm, conclude that more needs to be known about how this paradigm interacts with statistical learning and speech perception. Past results with this monitoring technique have pointed to an end-of-clause effect in parsing—a structural effect—but we here show that the issues are a bit more nuanced. Firstly, we report two Experiments (1a and 1b), which show that reaction times (RTs) are affected by two factors: (a) processing load, resulting in a tendency for RTs to decrease across a sentence, and (b) a perceptual effect which adds to this tendency and moreover helps neutralize differences between sentences with slightly different structures. These two factors are then successfully discriminated by registering event-related brain potentials (ERPs) during a monitoring task, with Experiment 2 establishing that the amplitudes of the N1 and P3 components—the first associated with temporal uncertainty, the second with processing load in dual tasks—correlate with RTs. Finally, Experiment 3 behaviorally segregates the two factors by placing the last tone at the end of sentences, activating a wrap-up

operation and thereby both disrupting the decreasing tendency and highlighting structural effects. Our overall results suggest that much care needs to be employed in designing click-detection tasks if structural effects are sought, and some of the now-classic data need to be reconsidered.

Keywords Tone monitoring · Processing load · Position effect · ERPs · Wrap-up

The Past

In a recent paper, Franco, Gaillard, Cleeremans and Destrebecqz (2015) applied the click-detection technique to a study of speech segmentation and statistical learning, concluding that whilst the click-detection task appears to be a promising way of assessing statistical learning, more needs to be known about both the underlying mechanisms of this type of learning and how the click detection interacts with these mechanisms and speech perception more generally. Here we provide a reassessment of the click-detection, or tone-monitoring, paradigm that connects to Franco et al.'s second point, especially as it concerns the study of parsing.

Click detection originated from the click-location paradigm, the latter an experimental technique first employed in the mid 1960s as a way of testing whether the clausal hierarchies postulated by generative grammarians reflected how people in fact conceptualize them in performance. As Fodor and Bever (1965) chronicle, this being the first study to employ click location in a parsing study, the click paradigm was at the time used by phoneticians as a means to probe perceptual consistency (p. 415), a phenomenon in which a processing unit can be found to resist interruption, hence constituting a “perceptual unit”. In particular, Fodor and Bever used the click-location technique to

✉ David J. Lobina
davidjames.lobina@ub.edu

Josep Demestre
josep.demestre@urv.cat

José E. García-Albea
jegarcia.albea@urv.cat

¹ Department of Philosophy/LOGOS,
University of Barcelona, Barcelona, Spain

² Department of Psychology, Universitat Rovira i Virgili,
Tarragona, Spain

determine if there were any such perceptual units in parsing, an issue and a version of the paradigm that will not feature much here. The monitoring version of the click paradigm was being developed at around the time the location version was at its peak of popularity, and it is the technique we will employ here.

In general, the click paradigm consists of superimposing a short, extraneous sound—a click, a tone, or else—over some linguistic material, which is then played to subjects over headphones. In the location version of Fodor and Bever (1965), the participants would be asked to write down the sentence they had just heard in order to then mark where they thought the click was placed. It was not a matter of evaluating the participants' accuracy in the task—they were indeed very inaccurate—it was instead an endeavor to map the errors subjects make, so that a comparison could be drawn between the objective position of the click and the position in which participants subjectively perceive it.

Fodor and Bever (1965) reported that even though participants had a tendency to perceive a click before its objective position (i.e., a left bias; p. 419), the overall majority of clicks, as subjectively perceived, were displaced towards clausal boundaries. Thus, in a sentence such as (1), clicks placed either between *was* and *happy* or between *was* and *evident*, marked below with the | symbol, would be perceived at the main clause boundary, that is, between *happy* and the *was* following it, this boundary marked by the || symbol.

- (1) That he was | happy || was | evident from the way he smiled

A biclausal sentence of these characteristics exhibits a certain complexity, as it contains various internal phrases and boundaries, but the results reported in Garrett, Bever, and Fodor (1966) suggest that clicks only ever migrate to the deepest constituent boundary—that is, between clauses. Similarly, Bever et al. (1969) concluded that within-clause boundaries do not appear to affect segmentation strategies in the location version. Put together, these results were taken as evidence that the clause is an important unit of processing, perhaps even constituting a perceptual unit (Fodor, Bever & Garrett, 1974). Furthermore, the clause-by-clause process seems to be solely the effect of syntactic properties, as other factors were controlled for and did not seem to affect the results (amongst others, pitch, intonation, a response bias, memory recall, etc.; see Garrett et al., 1966 and Bever, 1973 for details).

The last point was contested by Reber and Anderson (1970); by employing much simpler sentences (monoclausals such as *open roadside markets display tasty products*), they argued that the evidence suggested that (a) a right bias was actually operative, and (b) extra-linguistic factors were in fact responsible for the errors participants made. They also reported a tendency for subjects to mislocate

some clicks to the boundary between the subject and the verb, which might suggest that this break is also important for the processor, something that was explicitly denied in Fodor et al. (1974, p. 336).¹

These publications employed the location version of the click paradigm, an offline experimental technique that is currently largely out of favor (see Levelt, 1978, for discussion of its flaws). Abrams and Bever (1969) developed a detection or monitoring version, an online technique that is much more reliable, and the one under the microscope here. In this version of the paradigm, participants are required to press a button as soon as they hear the click or tone, and thus the analysis centers on the fluctuations in reaction times (RTs). The idea is that processing a sentence and monitoring for a tone compete for attentional resources, making tone monitoring a dual task. In general, the contention with this version of the paradigm is that reaction times ought to be greater at locations that require more working memory resources; that is, a correlation between reaction times and structural complexity.

The first studies employing this version of the paradigm also used biclausal sentences, with Abrams and Bever (1969) finding that clicks before the major break were reacted to more slowly than clicks *at* the major break or just *after* it in sentences such as *in addition to his wives || the prince brought the court's only dwarf* (the major break is once again marked by the || symbol). Similarly, Holmes and Forster (1970) found that RTs in the first half of a biclausal sentence were greater than in the second half. This phenomenon has been termed the end-of-clause effect (Bever & Hurtig, 1975) and constitutes a clear structural effect: the end of a major clause seems to involve some special processing.

Surprisingly, click detection has not been employed as much as it perhaps deserves, given its apparent sensitivity to the different cognitive loads the parser goes through within and between clauses in complex sentences (see Levelt, 1978, this time for discussion of how tone monitoring does not share the flaws of the location version of the paradigm). After Flores d'Arcais (1978) successfully used it to show that main clauses are usually easier to process than subordinates (and that the main/subordinate order exerts less memory resources than the subordinate/main order), the 1980s and 1990s hardly exhibit any other study employing this technique. Unsurprisingly, Cohen and Mehler (1996) considered their work a “revisit” to the paradigm—making our study a further visit—when they reported a number of clearly structural effects: RTs to tones at the boundary of reversible object relatives were greater than at structurally

¹Chapin, Smith and Abrahamson (1972) reported a similar result with biclausal sentences, but there are some problems with the materials they employed (see Fodor et al., 1974, pp. 336 et seq. for relevant discussion).

identical subject relatives or in other positions of a normal object relative, and a similar effect was found with semantically reversible and irreversible sentences, with the former exhibiting higher latencies (we will come back to these data in the last section). Recently, this monitoring task has been successfully employed in a word segmentation study (Gómez, Bion & Mehler, 2011), the work Franco et al. targeted in their own paper, and structural effects are also reported in that study.

It is hoped that the results we report here are further evidence for the usefulness of tone monitoring in the study of language comprehension. In particular, we report the results of four experiments with the detection version of the click or tone paradigm, one in combination with a recording of event-related brain potentials (ERPs), which when put together demonstrate the following: (a) RTs are affected by two factors: (i) processing load, which results in a tendency of RTs to decrease across a sentence, the result of the flow of linguistic information the processor successively receives (incrementality), and (ii) a strong perceptual effect seemingly operative in all monitoring tasks (Experiments 1a and 1b); (b) these two factors (one psycholinguistic, the other perceptual) can be discriminated by taking a record of ERPs during the monitoring task (Experiment 2); (c) the P3 component recorded in the ERP experiment can be usefully employed to measure processing effort in dual tasks such as the one administered here; and (d) psycholinguistic and perceptual factors can be behaviorally segregated by placing tones at the end of sentences, thus triggering a wrap-up operation, disrupting the decreasing tendency and highlighting structural effects in doing so (Experiment 3). These data provide the appropriate background for both a re-analysis of past results obtained with this technique and a discussion of the paradigm's strengths and weaknesses, both of which we undertake throughout the paper.

Current concerns

Cutler and Norris (1979) offer a thorough discussion of three kinds of detection tasks employed in psycholinguistics and argue that phoneme- and word-monitoring tasks diverge from a tone-monitoring task in that, inter alia, the former exhibit a general decrease in RTs across a sentence, an effect Cutler and Norris deny to the tone monitoring experiments of Abrams and Bever (1969) and Holmes and Forster (1970). This conclusion, however, is based on a less-than-careful analysis of the data that Abrams and Bever (1969), in particular, report. As mentioned in the previous section, these authors established three different click positions in sentences such as *since she was free that | day | her | friends asked her to come (before the main clause break, in the clause break, and right after the clause break, all*

marked with |), and the RTs they obtained certainly exhibit a decrease: 243 ms., 230, and 216. So why do Cutler and Norris (1979) conclude otherwise?

Abrams and Bever (1969) exposed their participants to repeated presentations of the same material, and participants' performance progressively improved. The RTs we have just provided were those of the first presentation, and it is only in the next two presentations wherein the linear decrease in RTs from the first to the third position disappears—the pattern that Cutler and Norris (1979) focus on. Given that these participants were reacting to familiar sentences *and* click positions in the second and third presentations, those responses are not comparable to those of the other monitoring tasks Cutler and Norris (1979) discuss (or other click-detection studies). Cutler and Norris (1979, p. 129) may have been too quick to deny tone monitoring a decreasing tendency if their analysis is based on the successive presentations Abrams and Bever (1969) carried out. Instead, the decrease in RTs seems to be a clear pattern in tone monitoring, as confirmed in the data of Holmes and Forster (1970) and Bever and Hurtig (1975).²

Holmes and Forster (1970) broach the decreasing tendency when they suggest that their participants must have been experiencing “maximal uncertainty” at the beginning of a sentence, something that is plausibly reflected in the high RTs for tones placed in the first clause. This maximal uncertainty makes reference to the predictions the parser generates during processing, and thus constitutes a structural phenomenon—a psycholinguistic factor. Namely, and following Holmes and Forster (1970), the processing load towards the end of a clause ought to be minimal, given that ‘structural information conveyed by the last few words would tend to be highly predictable’ (p. 299). The latter is in principle entirely compatible with the end-of-clause effect, which refers to the load specifically involved in closing off syntactic nodes and clauses, but there is certainly a question as to how these two effects relate to each other.

It is important to note that monitoring for a tone during the processing of a sentence is a type of dual task. Parsing a string would constitute the primary task on account of it being unconscious, automatic, fast, etc., meeting some of Fodor's (1983) criteria for modular processes, whilst monitoring and reacting to the tone would be the secondary task, a conscious and attentional phenomenon in competition for cognitive resources with the primary task (see Wickens,

²This is not to say that the memory resources involved in reacting to sentence-internal targets (such as those of phoneme- and word-monitoring tasks) are the same as those at work in targeting sentence-external elements (as in a click-detection task). Cutler and Norris (1979) certainly have other reasons to doubt this is the case, but we shall not discuss this issue here. Our point is merely that the decreasing tendency appears to be present in the three types of monitoring they discuss.

Kramer, Vanasse, & Donchin, 1983 for a discussion of dual tasks involving tone monitoring as secondary task; we will come back to this in much more detail below). In the particular case of Holmes & Foster's data (and, as we shall argue, those of Abrams & Bever), then, the explanation would be that the cognitive resources exerted by the primary task (parsing a string) are much greater at the beginning of a sentence, while later on the attentional mechanisms at work in the perception and monitoring of a tone have access to more resources—i.e., there is less competition for resources between the primary and the secondary task—and hence reactions to later tones ought to be faster.

Though Abrams and Bever (1969) and Holmes and Forster (1970) explain their data in terms of the processing load associated to the end of a clause, it is noteworthy that, in the case of Abrams and Bever (1969) at least, the click placed at the end of the major break also constitutes the end of a subordinate clause and the first of a series of three tones (one tone position per experimental sentence), and at that precise point the processor is in a state of great uncertainty indeed, for a significant amount of linguistic material is yet to come.³ Thus, the pattern reported in this study may not be the sole result of an end-of-clause effect, as this effect and the general tendency of RTs to decrease in monitoring tasks were not directly related, or controlled for, in these experiments.

In this study, we shall track the processing load of parsing monoclausal sentences rather carefully by focusing on two operations: phrase completion and verb-noun integration (subjects and objects). Taking our heed from the materials employed by Reber and Anderson (1970), we will use rather simple experimental sentences. That is, no biclausal sentences, ambiguous constructions, or relative clauses will feature in what follows; instead, unambiguous, declarative, active sentences will form the basis of our investigation. The plan is to employ three tone positions, one per sentence, and probe how the decreasing tendency relates to the processing of head-complement(s) phrases, the general geometry all syntactic phrases adhere to (as argued by Moro, 2008, following much earlier work on so-called X-bar theory; it is more accurate to call this geometry the specifier-head-complement, but the specifier position is usually empty in most phrases).

Monoclausal, subject-verb-object Spanish sentences were constructed for the purposes of this investigation. Starting from a matrix *proposition*—that is, a predicate and its arguments—two different types of sentences were

created. *Type A* sentences exhibited a complex subject but a simple object, while the reverse was the case for *Type B* sentences. By a complex subject or object is meant a noun phrase (composed of a determiner and a noun) which is modified by another noun phrase (also composed of a determiner and a noun, but introduced by a preposition). A simple subject or object, on the other hand, would simply be composed of a determiner and a noun. The following are the experimental sentences to be used in all the experiments reported in this paper, where the | symbol identifies the boundaries under study in the first three experiments.

Type A: El candidato | del partido | se preparó | el próximo discurso.

'The party's candidate prepared his next speech'.

Type B: El candidato | ha preparado | un discurso | sobre la sanidad.

'The candidate has prepared a speech about the health service'.

The | symbol indicates the relevant boundaries, but this is not where the tone was actually placed in the first three experiments. As will be described in the methodology sections below, the tones were placed on the second syllable after the relevant boundary in order to make sure that the parser had seen the previous syllable fully by then, thus readjusting its operations. In particular, we wanted to ensure that the parser had completed the previous phrase and moved on, as those were the precise locations we were interested in (this is especially important for the predictions we describe in the following paragraph). Thus, we targeted locations where we could be certain that the parser had completed a phrase, allowing us to better predict cognitive load vis-à-vis noun-verb integration. In addition, the tone was always placed right after either a preposition or the verb's auxiliary and thus before the content words within PPs and VPs (before nouns and verbs, that is), thereby controlling for any sort of PP- or thematic integration the processor would carry out once it encounters the relevant content of VPs and PPs.

Our general hypothesis is that RTs will tend to decrease within each sentence type due to the corresponding decrease in psycholinguistic uncertainty, which follows from the incremental nature of parsing (for a review, see Harley, 2001). However, this decrease in RTs ought to apply differently across sentence type, yielding the following predictions, which in this case stem from the two parsing operations we are tracking—phrase completion and noun-verb integration—and the somewhat simple structure of our sentences. In the first position, the parser has processed the same material in type A and type B sentences, identifying the noun phrase *the candidate* as the subject of the sentence, following the canonical subject-verb-object(s) order

³It is harder to evaluate the results of Holmes and Forster (1970), as these authors also found that RTs in a clause break were shorter than in positions wherein the tone would interrupt a large number of constituents, but in some sentences the clause break would precede the no-clause break, while in others it was the other way around. We are also not provided with the exact RTs in this study.

in Spanish, and thereby predicting the appearance of the verb. Thus, the cognitive load should be equal and the RTs similar.⁴ In the second tone position, the verb prediction is borne out in type B sentences and the parser successfully closes the subject noun phrase, whereas in type A sentences the parser is completing a longer subject noun phrase (a more complex head-complement structure) and the verb prediction is still active. Moreover, in type B sentences, the parser has integrated the verb and the subject noun phrase and now expects an object noun phrase, whilst in type A sentences the parser is yet to conduct any integration. In this case, then, the cognitive load should be greater in type A sentences and RTs higher to those of type B sentences. Finally, in the third tone position, the parser has integrated subject and verb in type A sentences and now predicts an object noun phrase, whereas in type B sentences the parser has successfully integrated part of the object noun phrase (the main part of a complex head-complement structure). In this case, too, type A sentences should involve more cognitive load and therefore higher RTs at this position.

These predictions are well motivated. Syntactic operations are more prominent than semantic and contextual factors in simple, active sentences (Pickering & van Gompel, 2006; van Gompel & Pickering, 2009), and thus we expect the integration of verbs and nouns to be rather central, especially the appearance of the verb, a sentence's central element. The completion of the noun-verb integration ought to reduce the parser's uncertainty as it processes a sentence, and this ought to be reflected in the data in the form of the decreasing tendency of RTs within a sentence, modulo the across-sentence-type differences we have predicted. Having outlined the general approach, we now turn to the experimental data.

Experiment 1

We report two slightly different experiments in this section. Following past practice with the tone-monitoring technique, we first report an experiment that only makes use of experimental sentences (Experiment 1a). This is then followed by an experiment that in addition contains filler sentences and a comprehension task (Experiment 1b). We decided to do this in order to evaluate our data in conditions similar to past experiments, first, and then compare this design with a more contemporary set-up.

⁴Note that what the parser has processed in this first position is one single syllable after the relevant boundary—the syllable where the tone is positioned is *being* processed—and this would not constitute enough information for the parser to identify this one syllable as either a preposition (in type A sentences) or an auxiliary (in type B sentences), and thus generate different predictions in each case.

Experiment 1a

Method

Participants Eighty-eight psychology students (20 male, 68 female) from the Rovira i Virgili University (Tarragona, Spain) participated in the experiment for course credit. The mean age was 20 years, and participants had no known hearing impairments. All were native speakers of Spanish.

Materials Two variants of monoclausal, active, declarative, subject-verb-object Spanish sentences were constructed from 60 matrix *propositions*. Type A sentences exhibited an [NP-[PP-NP]-[VP-NP]] pattern whereas type B sentences manifested a [NP-[VP-NP-[PP-NP]]] form—these are the structural conditions of the experiment. All sentences are unambiguous, composed of high- or very high frequency words, according to the corpora and classification in Almela, Cantos, Sánchez, Sarmiento and Almela (2005) (which was cross-checked with Sebastián-Gallés, Martí, Carreiras & Cuetos, 2000), and with a total length of 20 syllables. The sentences were recorded in stereo with a normal but subdued intonation by a native, male speaker of the Spanish language using the Praat software on a Windows-operated computer. Three tone positions per sentence were established, the three positional conditions of the experiment (1-2-3). Tones were placed on the vowel of the second syllable following the relevant boundary, so that the processor could use the first syllable (usually a preposition, the beginning of a preposition, or the auxiliary heading the verb) to “disambiguate” the location the parser was at that moment, thereby completing whatever phrase the parser was processing at each stage. The software Cool Edit Pro (Version 2.0, Syntrillium Software Corporation, Phoenix, AZ, USA) was employed to generate and superimpose tones with a frequency of 1000 Hz, a duration of 25 ms., and a peak amplitude equal to that of the most intense sound of the materials (80 dBs). Every sentence had one tone only, and in order to make sure that every item went through every condition, three different copies of each experimental item were created, totaling 360 experimental sentences. A further 12 practice items were created, two items per experimental condition.

Procedure The design of the experiment was a 2 (type of sentence factor) by 3 (tone position factor) within-subjects, within-items factorial, and therefore six lists were created. Each version was arranged according to a Latin square (blocking) method so that the items were randomized within and between blocks. Participants were randomly assigned to each list. The experiment was designed and run with the DMDX software (Forster & Forster, 2003) and administered in a sound-proof laboratory with low-to-normal illumination

in which a maximum of four subjects at a time would be tested. The sentences were presented over the headphones binaurally and participants were instructed to hold a key-pad with their dominant hand in order to press a button as soon as they heard the tone. They were told to be as quick as possible, but to avoid guessing. Once a sentence had finished, an instruction on the computer screen stated that the next sentence would be presented upon pressing the space bar, giving subjects control over the rate at which the sentences were presented. The experimental session consisted of 60 items and the DMDX software was used to measure and record reaction times. The whole session lasted around 20 min.

Results

The responses of eight subjects had to be eliminated for a variety of reasons. Six of these were due to technical problems with the coding of the computer programme and/or the equipment, while the other two did not meet reasonable expectations regarding average performance (one failed to register a single response).

The reaction times of the remaining 80 subjects were collected and trimmed with the DMDX programme. A response that occurred before the tone or 3 s after the tone was not recorded at all (in some cases, 3 s after the tone meant that the sentence had long finished), while responses deviating 2.0 SDs above or below the mean of each participant were eliminated (this affected 4.3% of the data). The resultant measures were then organized according to experimental condition. The analysis of reaction times was carried out with the SPSS package (IBM, USA). Table 1 collates the RTs per condition.

As can be observed in Table 1, RTs are greater in position 1 and decrease thereon for each sentence type. Moreover, RTs to type A sentences appear to be slightly higher than to type B sentences. A repeated-measures analysis of variance showed that the *tone position factor* was significant in both the subjects and items analyses ($F_1(2, 158) = 144$, $p < .001$, $\eta_p^2 = 0.647$; $F_2(2, 118) = 295$, $p < .001$, $\eta_p^2 = 0.834$; $\min F'(2, 265) = 96.76$, $p < .001$), while the *sentence type factor* was only significant in the subjects analysis ($F_1(1, 79) = 4.66$, $p < .05$, $\eta_p^2 = 0.056$;

$F_2(1, 59) = 2.48$, n.s.; $\min F'(1, 114) = 1.61$, n.s.). There was no interaction between the two experimental factors (all $F_s < 1$).

Pair comparisons between the three positions of the *tone position factor* showed that the differences in RTs were all significant: 1–2 ($t_1(79) = 10.9$, $p < .01$; $t_2(59) = 15.8$, $p < .01$); 1–3 ($t_1(79) = 13.5$, $p < .01$; $t_2(59) = 24.1$, $p < .01$); 2–3 ($t_1(79) = 7.9$, $p < .01$; $t_2(59) = 7.0$, $p < .01$).

Experiment 1b

Method

Participants Seventy-seven psychology students (eight male, 69 female) participated in the experiment for course credit. This was a different set of participants from Experiment 1a. The mean age was 22 years, and no subject had any known hearing impairment. All were native speakers of Spanish.

Materials The experimental items were the same as in the previous experiment. Sixty new sentences were now constructed to act as fillers. Twenty-four of these fillers were long, biclausal sentences, 24 were monoclausal sentences with a different word order from the canonical subject-verb-object, and the remaining 12 fillers were exactly like the experimental items but did not carry a tone. Twelve other fillers did not carry a tone, either; in total, 20% of the items did not have a tone. Regarding the comprehension task, 24 questions were constructed, 12 for the fillers, and 12 for the experimental items. The questions were rather simple in formulation and would query an uncomplicated aspect of either the subject, the object, or the verb of the corresponding items. The answer required was either a *yes* or a *no*. All other significant aspects of the task (generation and introduction of tones, etc.) remained unchanged from the previous experiment.

Procedure The same as in the previous experiment, but with the addition of the fillers and the comprehension task. The fillers and the experimental sentences were randomized together for this version, which naturally included the questions some of these items were associated with. Regarding the comprehension task, each question appeared on the computer screen and the participants recorded their answers by pressing either the S key (for *sí*, that is, yes) or the N key (for *no*). The overall task was divided into three even blocks. During the break, the computer screen would turn white and subjects would be instructed to rest and relax, but to not disturb the others. The break would last two minutes, and at the end the screen would turn black in order to signal that the break had finished. A third and final white screen indicated that the overall session had finished. In all

Table 1 Experiment 1a. RTs per tone position per sentence type (mean RT with standard deviations in parentheses)

Sentence type	Tone position		
	1	2	3
A	257.22 (59.1)	222.51 (41.0)	206.78 (40.1)
B	252.40 (52.0)	217.33 (43.9)	205.26 (44.3)

other significant respects, the new task remained exactly the same as in the previous experiment. The session was now significantly longer taking close to 40 min to complete.

Results

Ten participants were eliminated as they did not meet reasonable expectations regarding average performance. In particular, two subjects had an average response time that was close to 2 s, while another subject failed to record a single response. An analysis of the comprehension task showed that participants hardly made any errors, and apart from a participant who erred in 40% of the questions, everyone else was well under that figure. As we had settled on a 30% cut-off, only this subject was eliminated. The responses of the remaining participants were collected and trimmed following the same procedure of the previous experiment. Again, responses deviating 2.0 SD above or below the mean of each participant were eliminated; in this case, 4.0% of the data were affected. As before, the reaction times, summarized in Table 2, were analyzed with SPSS.

As in Experiment 1a, and for each sentence type, RTs were greatest in the first tone position and decreased thereon. The analyses of variance with subjects and items as random factors once again showed that the *tone position factor* was significant ($F_1(2, 130) = 70.21, p < .001, n_p^2 = 0.519; F_2(2, 118) = 36.61, p < .001, n_p^2 = 0.383; \text{min} F'(2, 218) = 23.77, p < .001$), while the *sentence type factor* did not prove to be significant in either analysis ($F_s < 1$). The interaction effect was also not significant ($F_1(2, 130) = 1.5, \text{n.s.}; F_2 < 1$).

Regarding the significance of the tone position treated as a simple effect, the 1-2 and 1-3 pairs proved to be significant (1-2: $t_1(65) = 10, p < .01; t_2(59) = 6.5, p < .01$; 1-3: $t_1(65) = 9.6, p < .01; t_2(59) = 7.7, p < .01$), but this was not the case for the 2-3 pair ($t_1(65) = 1.5, \text{n.s.}; t_2(59) = 1.0, \text{n.s.}$).

Discussion

As is clear from the results listed above, the decreasing tendency in RTs was confirmed, the tone position factor was significant in every analysis (this was not the case for the

sentence type factor), but there was no interaction. Thus, whilst our general hypothesis was supported, many of the more specific predictions were not confirmed. The decreasing progression of the RTs is rather robust, and the high significance of the (tone) position factor is further confirmation. This is of course in line with the expectation that processing load decreases as the sentence is presented—the least linguistic material to process, the easier it will be to respond to the tone, following from incrementality.

This cannot be the whole story, however, as there were no differences in RTs across sentence type, which differed in terms of structure, and this is surprising. Namely, each tone was placed in a rather different segment in each sentence type, and thus the parser cannot be computing the same predictions at each tone position (except for the first tone position)—i.e., the parser's uncertainty cannot be the same. This ought to be especially significant when it comes to integrating verbs and nouns during the course of a sentence, but as the data show the earlier or later appearance of the verb and whether noun phrases were simple or complex does not appear to have had much of an effect. Thus, our data cannot be entirely explained in terms of processing load. Interestingly, this is also the case in Abrams and Bever (1969), as these authors used biclausal, complex sentences and the course of incremental parsing in that study ought to have been different to what we obtained here—and yet they also reported a decreasing tendency in RTs across three tone positions.

We postulate that there must be a perceptual factor at play in monitoring tasks; roughly stated, the later the tone appears, the more prepared the participants are to respond to it. This would accord well with Cutler and Norris's (1979) own analysis of phoneme- and word-monitoring, as processing load on its own can certainly not explain the decreasing tendency in those monitoring tasks, lending some credence to the analogy we drew between the three monitoring tasks in the previous section. If this is the case, there would be two types of uncertainties to track in monitoring tasks: one psycholinguistic, stemming from incrementality—viz., what linguistic material is there left to process?—the other perceptual—viz., when will the tone appear?—which we shall call the *position* effect. On the one hand, then, participants would be progressively better prepared to respond to a tone the more settled they are during the experiment. This would be a matter of how attentional mechanisms function in such tasks, and there is evidence that these considerations apply regardless of the type of input in which the tones are placed (Wickens et al., 1983; Sirevaag et al., 1993). On the other hand, though, as the sentence is being presented a participant would increasingly be less surprised/uncertain when the tone finally appears, and therefore participants ought to be faster in responding to the tone when it appears towards the latter part of sentences.

Table 2 Experiment 1b. RTs per tone position per sentence type (mean RT with standard deviations in parentheses)

Sentence type	Tone position		
	1	2	3
A	340.71 (89.8)	290.86 (79.1)	283.00 (67.4)
B	335.42 (88.9)	296.54 (96.5)	291.25 (81.5)

As such, the results of our first two experiments—a decrease in RTs and no interaction between experimental factors—would be the product of the joint effect of perceptual and psycholinguistic factors. We are not conflating the processing load involved in parsing a sentence and the perceptual factor we are postulating into one single factor. Rather, what we are suggesting is that our data are the result of the combination of *both* factors, which may or may not be separately observed and may or may not be behaviorally segregated (we evaluate these two possibilities in Experiments 2 and 3). If this conjecture is correct, then the greater RTs in the first tone position in Abrams and Bever (1969) may not have been solely due to an end-of-clause effect, but the result of the combination of perceptual and psycholinguistic factors. Indeed, given that past studies did not consider any perceptual factors and thus did not control for tone position, we are unsure as to whether the end-of-clause effect is all that well supported.

That being so, the results reported in Flores d'Arcais (1978) and Cohen and Mehler (1996) are clearly structural rather than perceptual, and as such tone monitoring must be sensitive to both factors. Our own results yielded structural factors too other than our (confirmed) general contention regarding processing load and RTs, at least in the subjects analysis of Experiment 1a, where the sentence type factor proved to be significant. However, the fact that this factor was only significant in this analysis and there was no interaction with the tone position factor needs to be explained. We shall elaborate a possible explanation as we proceed but advance now that both our materials and the tone positions we employed were significantly simpler than in previous experiments, and as such the perceptual factor we have identified may have abated the structural differences somewhat, thus producing less clear structural effects (this will be evaluated properly in Experiment 2). We should like to emphasize that the effects are “less clear” rather than absent, as we did obtain a general structural effect: the decrease in processing load *is* a parsing, and thus a structural, phenomenon. What we did not find was structural differences between two types of sentences that differed along an admittedly rather similar dimension (shorter or longer subject or object noun phrases), but we should not conclude from this that tone monitoring is not sensitive to structural factors or that the sentences were not appropriately processed in our experiments. A corollary of this point is that it ought to be possible to manipulate either the materials or the tone positions of Experiments 1a/1b in order to unearth clearer structural effects, and we shall undertake this in Experiment 3.

Two further aspects of our data are worth discussing. First, the general decreasing tendency already observed in the now-classic results of the 1960s and 70s, as described earlier, receives further confirmation by an analysis we

conducted on the RTs to the filler sentences from Experiment 1b. Given that the tones were introduced in a somewhat random manner in the construction of the fillers, a correlation analysis was conducted in which x stood for the number of syllables after which the tone would appear in each item and y was the reaction time to the tone. The Pearson's correlation was $r_{xy} = -.633$, $p < .01$, indicating that the greater the number of syllables (that is, the deeper into the sentence the tone is), the lower the reaction time to it.

Secondly, it is also of some interest that we obtained a slightly different level of performance in Experiment 1b compared to Experiment 1a: the RTs in the former are significantly higher than in the latter (313.67 ms. for 240.68) and the statistical results are also slightly different. Indeed, the 2-3 and 2-3 pairs proved to be not significant in the statistical analyses of Experiment 1b. This is somewhat unexpected, for Experiment 1b included fillers and a comprehension task and this design was expected to highlight the structural differences. That is, the inclusion of fillers and a comprehension task ought to have resulted in clearer structural effects, but these were in fact more pronounced in Experiment 1a (and in past studies without filler sentences and comprehension tasks, we note). As mentioned, this may have been the result of the relative simplicity of our materials, in general, and the locations of the tones, in particular, a point we will further expand throughout the paper.

Moving on, in the next experiment we switch tack and proceed to attempt to discriminate the psycholinguistic and perceptual factors we have postulated but which are not easily discernible in the behavioral data so far obtained: processing load and the position effect. In order to do so, we can combine tone monitoring with the recording of electrophysiological responses to external stimuli (i.e., ERPs), which will allow us to track two different ERP components, one related to processing load (and linguistic uncertainty), the other to the position effect (and temporal uncertainty). If there is a correlation between these ERP waves and RTs, our interpretation of the data we obtained in Experiments 1a and 1b would be confirmed.

Experiment 2

In this experiment, only type A sentences from the previous experiments were employed. As we aimed to discriminate the position effect from the processing load the parser goes through as the sentence is presented, there really was no reason to employ both sentence types; the tone positions, however, remain the same.

We decided to concentrate on two ERP components, yielding two broad but relevant predictions. It was, first of all, hypothesized that the N1 wave, a component associated with temporal uncertainty (Näätänen & Picton, 1987),

would correlate with the RTs, and thus its amplitude would be highest at the first tone position, the perceptual uncertainty of the participants being greatest at that point, and decrease thereon. This part of the experiment aimed to evaluate the significance of the position effect, and the N1 is a pertinent component for such a task, given that it tracks perceptual processes rather than (higher) cognitive ones.⁵

The second component of interest is the P3 (or P300), a component whose amplitude to a secondary task has been shown to be affected by the difficulty of the primary task in dual-task settings such as ours. Past results with dual-task experiments (e.g., Wickens et al., 1983) indicate that the P3 associated with a secondary task (in this case, reacting to the tone) will have a low amplitude if the primary task (here, parsing the sentence) is of considerable difficulty. In other words, there will be a negative correlation between the fluctuations in difficulty in a primary task and the amplitude of the P3 to a secondary task. In our experiment, as the primary task decreases in difficulty (as manifested by the linear decrease in RTs from the first to the third position), the amplitude of the P3 was predicted to increase from position 1 onwards. That is, as the sentence is being processed, the number of predictions the parser needs to fulfill is reduced, and thereby more resources can be allocated to responding to the tone, something that should be reflected in the amplitude of the P3. If this prediction is confirmed, it would justify long-held assumptions regarding the internal structure of dual tasks such as tone monitoring.⁶

Crucially for our purposes, the biphasic pattern we are hypothesizing is well established in the dual tasks literature. Both Sirevaag et al. (1993) and Wickens et al. (1983) report an N1-P3 pattern when an auditory probe is employed, and this is precisely what we are after: an N1 wave tracking perceptual processes and a P3 component tracking cognitive processes. In particular, we expect to obtain an N1 wave with a frontal distribution and a P3 with a more posterior-parietal distribution, thus singling out two independent components and, in the case of the P3, ruling out a novelty P3 and instead identifying a wave representing a distribution of processing capacity between concurrent tasks (Giraudet, St-Louis, Scannella & Causse, 2014; Käthner, Wriessneger, Müller-Putz & Kübler, 2015). If these two waves turn out to be present in the data, and their amplitudes go in the direction we are postulating, we would have clear

evidence for the two factors we have postulated: one perceptual, one linguistic. To our knowledge, moreover, this is the first time that the P3 is employed in a study of syntactic processing as a metric of processing load, and we hope our results constitute evidence for its general usefulness in psycholinguistics. Naturally, these two hypotheses hold if and only if the pattern in RTs obtained in the previous experiments does not vary, and we hypothesized that this would be the case indeed.

Method

Participants Eighteen psychology students (two male, 16 female) from a pool of different courses participated in the experiment. This was a different set of participants from the previous two experiments. The mean age was 22 years, and subjects had no known hearing impairments. All were native speakers of Spanish.

Materials The same as type A sentences from the previous experiments, but these now numbered 120 items. There were no fillers, nor a comprehension task.

Procedure Participants were seated in a comfortable chair in a sound-attenuated, darkened, and dimly illuminated room. One individual was tested at each time. In this experiment, they were exposed to a total of 120 items, presented in three blocks. The pauses were of the same length as in Experiment 1b, and apart from the EEG measures that were undertaken, the task remained the same as in the previous experiments. The EEG was recorded continuously by 19 Ag/AgCl electrodes which were fixed on the scalp by means of an elastic cap (Electrocap International, Eaton, OH, USA) positioned in accordance with the 10-20 International system. ERPs were algebraically re-referenced to linked earlobes offline. Electrode impedances were kept below 5 k Ω . All EEG and EOG channels were amplified using a NuAmps Amplifier (Compumedics, Charlotte, NC, USA) and recorded continuously with a bandpass from 0.01 to 30 Hz and digitized with a 2-ms. resolution. The EEG was refiltered off-line with a 25-Hz, low-pass, zero-phase shift digital filter. Automatic and manual rejections were carried out to exclude periods containing movement or technical artifacts (the automatic EOG rejection criterion was $\pm 50 \mu V$).

Results

Behavioral data

The reaction times of the 18 participants were collected and trimmed with the DMDX programme. As before, responses

⁵As an aside, we note that Näätänen and Picton (1987) chronicle a number of studies that also focus on the N1 wave and the monitoring of tones, but not in combination with linguistic material.

⁶There has been some discussion in the literature as to whether the P3 constitutes a good metric for processing load, but doubts seem to be stronger regarding single task experiments; dual tasks appear to provide the right setting for the employment of the P3 as a reliable measure of processing load (see Kok, 2001, for extensive discussion of these issues).

deviating 2.0 SDs above or below the mean of each participant were eliminated, which in this case affected 3.6% of the data. The final data is shown in Table 3.

As expected, the RTs manifest the exact same pattern as in Experiments 1a and 1b: reaction times decrease from the first position onwards. A repeated-measures analysis of variance showed that the *tone position factor* was significant for both the subjects and items analyses ($F_1(2, 34) = 39$, $p < .001$, $\eta_p^2 = 0.698$; $F_2(2, 238) = 93$, $p < .001$, $\eta_p^2 = 0.441$; $\min F'(2, 67) = 27$, $p < .001$). Regarding pair comparisons between the different tone positions (1 vs. 2, etc.), the analyses showed that all comparisons were significant: 1-2 ($t_1(17) = 6.2$, $p < .01$; $t_2(119) = 10.3$, $p < .01$); 1-3 ($t_1(17) = 6.6$, $p < .01$; $t_2(119) = 12.1$, $p < .01$); 2-3 ($t_1(17) = 3.8$, $p < .01$; $t_2(119) = 3.3$, $p < .01$).

Electrophysiological data

The data were processed using BrainVision Analyzer 2 (Brain Products, Gilching, Germany). Average ERPs were calculated per condition and per participant from −100 to 500 ms. relative to the onset of the tone, and before grand-averages were computed over all participants. A 100-ms. pre-tone period was used as the baseline. Only trials without muscle artifact or eye movement/blink activity were included in the averaging process. The analyses were based on 15 channels divided into five separate parasagittal columns along the anteroposterior axis of the head. The columnar approach to analyzing the ERP data provides both an anterior-to-posterior as well as a left/right comparison of ERP effects. The electrodes in each of two pairs of lateral columns (inner column: F3/F4, C3/C4, P3/P4; outer column: F7/F8, T3/T4, T5/T6) and on the midline column (Fz, Cz, Pz) were analyzed with three separate ANOVAs. The analysis of the midline column included the position factor (position 1 vs. position 2 vs. position 3) and the location factor with three levels (Fz vs. Cz vs. Pz). The analyses of the two pairs of lateral columns involved repeated measures ANOVAs with within-participants factors position (position 1 vs. position 2 vs. position 3), location (anterior, central, and posterior) and hemisphere (left and right).

Table 3 RTs per tone position (mean RT with standard deviations in parentheses)

Tone position		
1	2	3
325.05 (64.7)	266.53 (36.5)	247.60 (31.0)

Omnibus ANOVAs were followed up with pairwise comparisons intended to discern whether there were differences among the three tone positions. All post-hoc analyses were Bonferroni corrected. Based on prior reports, two time windows were selected for analysis of the mean amplitudes of the components of interest: the N1 component was analyzed from 120 ms. to 200 ms., and the P300 component was evaluated from 230 to 400 ms. The Greenhouse and Geisser (1959) correction was applied to all repeated measures having more than one degree of freedom in the numerator. In such cases, the corrected p value is reported. In order to not clutter the presentation of our results, we only report the main effect of the tone position factor and the significant interaction effects between this factor and the others.

Figure 1 depicts brain potential variations in the three midline electrodes included in the analyses. As can be observed, the three tone positions exhibit a clear biphasic pattern, with a first modulation in the N1 time window in frontal and central electrodes, followed by a second modulation in the P300 time window in the central and posterior electrodes.

N1 epoch (120–200 ms) During the N1 epoch, there was a main effect of position—midline column: $F(2, 34) = 24.021$, $p < .001$, $\eta_p^2 = 0.586$; inner column: $F(2, 34) = 14.939$, $p < .001$, $\eta_p^2 = 0.468$; and outer column: $F(2, 34) = 19.402$, $p < .001$, $\eta_p^2 = 0.533$. Bonferroni corrected pairwise comparisons showed that all three positions differ from each other significantly in the three columns (all $ps < .05$), reflecting a more negative-going amplitude for position 1 relative to position 2, and a more negative-going amplitude for position 2 relative to position 3. There was also a significant interaction between position and location in the midline column, $F(4, 68) = 5.046$, $p = .011$, $\eta_p^2 = 0.229$. Post-hoc comparisons in this column revealed that whereas in frontal and central electrodes position 1 was more negative relative to position 2, and position 2 more negative relative to position 3 (all $ps < .05$), there were no differences between the three positions in the posterior electrodes (all $ps > .20$). The position $\times$ location interaction was also significant in the inner column, $F(4, 68) = 6.313$, $p = .002$, $\eta_p^2 = 0.271$. Post hoc comparisons in this column revealed that whereas in frontal and central electrodes position 1 was more negative than position 2, and position 2 more negative than position 3 (all $ps < .05$), there were no differences between the three positions in the posterior electrode (all $ps > .21$). This two-way interaction was also significant in the outer Column, $F(4, 68) = 10.715$, $p < .001$, $\eta_p^2 = 0.387$. Post hoc comparisons in this column revealed that whereas in frontal electrodes position 1 was more negative than position 2, and position 2 more negative than position 3 (all $ps < .05$), there were no

Fig. 1 ERP waveforms for the three tone positions shown from a 100-ms. before tone presentation to a 500-ms. post-tone presentation. The waveforms depict brain potential variations in the three midline electrodes included in the analyses. Negative voltage is plotted up

differences between the three positions in central and posterior electrodes (all $ps > .52$). The two-way interaction between position and hemisphere and the three-way interaction between position, location, and hemisphere did not reach significance in the inner column ($Fs < 1$). In the outer column, the position $\times$ hemisphere interaction reached significance, $F(4, 68) = 7.566$, $p = .003$, $n_p^2 = 0.308$. Post hoc comparisons revealed that whereas in the left hemisphere position 1 was more negative than position 2, and position 2 more negative than position 3 (all $ps < .05$), there were no differences between the three positions in the right hemisphere (all $ps > .30$). In the outer column, the three-way interaction between position, location, and hemisphere reached significance, $F(4, 68) = 11.229$, $p < .001$, $n_p^2 = 0.398$. Post hoc comparisons revealed that whereas in the left anterior electrode position 1 was more negative than position 2, and position 2 more negative than position 3 (all $ps < .05$), there were no differences between the three positions in all other electrodes (all $ps > .60$).

P300 epoch (230–400 ms) During the P300 epoch, there was a main effect of position—midline column: $F(2, 34) = 16.827$, $p < .001$, $n_p^2 = 0.497$; inner column: $F(2, 34) = 26.515$, $p < .001$, $n_p^2 = 0.609$; outer column: $F(2, 34) =$

26.002, $p < .001$, $n_p^2 = 0.605$. Bonferroni corrected pairwise comparisons in the three columns showed all three positions to differ from each other significantly (all $ps < .05$), reflecting a more positive-going amplitude for position 3 relative to position 2, and a more positive-going amplitude for position 2 relative to position 1. There was also a significant interaction between position and location in the midline column, $F(4, 68) = 5.118$, $p = .004$, $n_p^2 = 0.231$. Post hoc comparisons in this column revealed that whereas in central and posterior electrodes position 3 was more positive relative to position 2, and position 2 more positive relative to position 1 (all $ps < .05$), there were no differences between the three positions in the frontal electrode (all $ps > .11$). The position $\times$ location interaction was also significant in the inner column, $F(4, 68) = 12.326$, $p < .001$, $n_p^2 = 0.420$. Post hoc comparisons in this column revealed that whereas in central and posterior electrodes position 1 was more positive than position 2, and position 2 more positive than position 1 (all $ps < .05$), there were no differences between the three positions in the frontal electrodes (all $ps > .12$). This two-way interaction was also significant in the outer column, $F(4, 68) = 3.967$, $p = .044$, $n_p^2 = 0.189$. Post hoc comparisons in this column revealed that whereas in central and posterior electrodes position 3

was more positive than position 2, and position 2 more positive than position 1 (all $ps < .05$), there were no differences between the three positions in frontal electrodes (all $ps > .31$). The two-way interaction between position and hemisphere and the three-way interaction between position, location, and hemisphere did not reach significance in the inner column ($ps > .28$) or at the outer column ($ps > .22$).

Discussion

As the behavioral data show, the prediction regarding the RTs pattern was confirmed; that is, RTs to the first tone are slowest, and then become faster thereon. Further, all pair comparisons were significant. Together, this allows us to discuss the ERP data in the terms we had devised.

The ERP data we obtained confirm both the hypothesized distributions of the N1 and P3 components and their amplitudes, as shown in Fig. 1. The N1 pattern indicates that participants are indeed uncertain as to when the tone is going to appear, and their uncertainty decreases as the sentence is being presented. This is of course unsurprising; as the sentence progresses and gets closer to the end, the chances that the tone will finally appear increase, and the uncertainty thereby decreases. We had already stated that the linear decrease in RTs was due to a combination of two factors and the N1 data confirm that there is indeed a purely perceptual factor at play, what we called earlier the position effect. Regarding the P3, its pattern can be explained in terms of task difficulty. As the amplitude of the P3 increases from position 1 onwards, and there is furthermore a negative correlation between RTs and the amplitude of the P3, this confirms that as the sentence is being processed the parser's unfulfilled predictions decrease, and thereby more resources can be allocated to monitoring the tone.

The biphasic pattern we have recorded confirms our analyses. First, the correlation between the amplitude of the N1 component and tone position confirms that there is a strong perceptual factor and that it has an effect on performance. Second, the correlation between the amplitude of the P3 and tone position confirms two interrelated points: (a) that the click detection is a dual task in which sentence processing is the primary task and tone monitoring the secondary; and, consequently, (b) that the fluctuations in processing load are in part due to the decreasing uncertainty the parser experiences, and thus dismissing alternative explanations in terms of response strategies, guessing the position of the tone, or the like.

The parser's decreasing uncertainty is certainly true of our materials and has received much confirmation in other contexts. However, it ought to be perfectly possible to construct materials in which the parser's uncertainty increases instead of decreasing, and this would be reflected in the RTs

(see ft. 7 below)—and in the amplitude of the P3. Having said that, we have also shown that linguistic uncertainty and incrementality interact with perceptual uncertainty in our task, as shown in the N1 amplitude, and previous studies employing the click-detection paradigm did not consider this particular factor. All in all, we have succeeded in discriminating—that is, recording—the two factors we had posited. In the next experiment, we shall show how they can in addition be behaviorally segregated to some extent, thereby highlighting structural effects with the tone-monitoring technique.

Experiment 3

A number of factors and some experimental evidence suggest that the end of a sentence ought to exert a particular cognitive load in the parser. We are not referring to the end-of-clause effect reported in Abrams and Bever (1969) and Bever and Hurtig (1975); we are not entirely sure about those data, given what we have said about the two factors that seem to be operative in monitoring tasks—and, in any case, in those studies the end of the clause was never the end of a sentence. Rather, what we have in mind is the self-paced reading literature, wherein a “wrap-up” effect has been reported, as evidenced in the tendency of reading times to be much higher towards the end of a sentence (Just, Carpenter & Woolley, 1982). This is *prima facie* a very natural result to obtain; it would be predicted to follow from the very general point that the more material the parser is inputted, the more strained working memory is likely to be. After all, the ceaseless inflow of material should result in a ever-greater number of open syntactic nodes, and these would only be “branched off” at the end of a sentence. Ambiguity resolution, the completion of the overall sentence, and the construction of the underlying proposition are some of the other phenomena that further suggest a wrap-up operation at this location.

The wrap-up operation was not applicable in the previous experiments, as the last tone position was nowhere near the end of any of the sentences. Here, we modify type B sentences from Experiments 1a/1b by changing the positions of the tone in order to probe if by placing a tone at the end of a sentence the strong tendency for RTs to decrease is disrupted. We only use type B sentences because these exhibit a complex noun phrase in the object position, and this is a better configuration for the purposes at hand, considering that the object position is towards the end of these sentences. Further, only one sentence type is necessary for our purposes, as we are only aiming to disrupt the decrease in RTs within a sentence—that is, we are interested in comparing tone positions within a sentence type, not across sentence types.

Three tone positions are maintained, but their locations were changed: one was placed at the beginning of the sentence and two within the verb's complex object (the new tone positions are shown in the following section). It was hypothesized that the wrap-up effect would be indeed applicable at the end of a sentence and therefore that the pattern in RTs should be different from the pattern observed in the previous experiments. In particular, we expected a V-shape pattern in which RTs to the first position were highest, descending significantly for the second position, but then raising for the third and last position, the postulated locus of the wrap-up operation. The aim of this experiment, therefore, was to highlight clearer structural effects with the tone monitoring technique.

Method

Participants Forty-eight psychology students (five male, 43 female) from the Rovira i Virgili University (Tarragona, Spain) participated in the experiment for course credit. This was a different set of participants from all other experiments. The mean age was 22 years, and none of the subjects had any known hearing impairments. All were native speakers of Spanish.

Materials Type B sentences from Experiments 1a and 1b were employed, but 14 of these had to be either slightly modified or substituted by new items. While preparing the experiment, it was noticed that in 14 of the 60 experimental sentences the parser could well carry out a wrap-up at the second boundary, as none of these sentences required any further material at that point (intonational phrasing notwithstanding). Consequently, it was decided that these 14 sentences be modified or replaced by new ones so that the experimental materials were as homogeneous as possible. The tone positions were also modified in order to evaluate the wrap-up effect, and as before, were always placed on vowels. The following sentence shows the new tone positions (where |, in this case, marks the placement of the tone):

- (2) El candi|dato ha preparado un di|scurso sobre la sani|dad.

The fillers were the same as in Experiment 1b, but in all other respects the task did not change.

Procedure The same as in Experiment 1a. The session took around 25 min to complete.

Results

The responses of 11 participants had to be eliminated for a number of reasons: some failed to achieve a reasonable

Table 4 RTs per tone position (mean RT with standard deviations in parentheses)

Tone position		
1	2	3
414.16 (51.3)	351.88 (33.9)	365.45 (37.4)

standard in performance, while others did not manage to register a single response. The reaction times of the remaining 37 subjects were collected and trimmed with the DMDX programme. As ever, responses deviating 2.0 SDs above or below the mean of each participant were eliminated, here affecting 3.8% of the data. The analysis of the reaction times, shown in Table 4, was once more carried out with the SPSS package.

In this experiment, RTs were also greatest in the first position, but there was no decrease from the second to the third position; instead, there was a slight increase. A repeated-measures analysis of variance showed that the *tone position factor* was significant in both the subjects and items analyses ($F_1(2, 72) = 98, p < .001$; $F_2(2, 118) = 110, p < .001$; $\min F'(2, 173) = 51.82, p < .001$). All post hoc pairwise comparisons proved to be significant: 1-2 ($t_1(36) = 12.5, p < .01$; $t_2(59) = 13.1, p < .01$); 1-3 ($t_1(36) = 9.0, p < .01$; $t_2(59) = 11.1, p < .01$); 2-3 ($t_1(36) = -3.9, p < .01$; $t_2(59) = -3.3, p < .01$).

Discussion

As predicted, the wrap-up effect was detectable with the tone-monitoring task, thereby disrupting the linear decrease in RTs, as can be seen in Fig. 2.

Fig. 2 RTs progression in Experiment 3

Indeed, even though RTs to the first position were greatest and there was a noticeable decrease from the first to the second position, the processing load associated with the wrap-up effect resulted in an increase in RTs from the second to the third position, in clear contrast with what was obtained in the previous experiments, and resulting in the V-shape pattern observed in Fig. 2. This would seem to indicate that a design can be found so that structural properties are brought out more clearly, resulting in the clear segregation of the two factors that have animated the whole discussion. Psycholinguistic and perceptual factors *can* be separated, vindicating the usefulness of tone monitoring in the study of parsing.

Note, however, that RTs to the first position are still significantly greater than RTs to the last position, indicating that the processor is more strained at the beginning than at the end of sentences. Also, we are not at all certain as to what sort of operations precisely the wrap-up involves, only that something peculiar to that location does take place. It seems safe to suppose, given incrementality, that what does not take place is the branching off of all the nodes the parser would have reputedly opened during sentence processing—the parser is supposed to complete phrases as it encounters them. This aspect of the wrap-up possibly merits more attention, but we here simply and tentatively suggest that what takes place is some sort of “putting it all together” process wherein the underlying proposition a sentence is said to express is completed and/or refined—or said otherwise, the overall syntactic object is finally completed.

Whether this effect can be related to the end-of-clause effect apparently unearthed in previous experiments is not so clear. In those studies, and as already stated, the end of a clause was in fact the end of a subordinate clause within a complex, biclausal sentence, with all the orbiting issues that arise therefrom. Moreover, the end-of-clause position was usually also the first tone position of the sentences employed in those studies, pointing to the probable role of the position effect. We come back to some of these issues in the last section, where we re-analyze some of the data obtained with the paradigm in past years and discuss the methodological issues to take into consideration when employing the tone-monitoring technique in the study of parsing.

Part of the past and something for the future

This is what we have done here. First off, we employed two types of simple, declarative sentences to probe if the linear decrease in RTs observed in phoneme- and word-monitoring tasks was also a factor—and if so, how strong of a factor—in the monitoring of a tone during the parsing of a sentence. In order to mitigate the probable decrease in RTs, the two types of sentences we constructed varied

in terms of whether complex noun phrases (i.e., complex head-complement structures) appeared in either the subject or object position (along with their relationship with the verb), the locations where the tones would be placed. The results of Experiments 1a and 1b showed two things: (a) a pronounced decrease in RTs for each individual sentence; and (b) no significant differences in the RTs to the same tone locations across sentence type. We postulated that these data were the result of the additive effects of two factors: (1) processing load, which decreases as the sentence is presented, thus releasing more cognitive resources to monitor the tone in doing so; and (2) a position effect, a purely perceptual factor involved in all monitoring tasks (as mentioned earlier, processing load cannot account for the decreasing tendency in other types of monitoring). These two factors were then separately observed in an ERP experiment (Experiment 2), and indeed separated in behavior in a modified version of the main experiment (Experiment 3).

The position effect seems to have gone entirely unnoticed in all previous studies of tone monitoring, whilst the processing load factor (which we believe to follow from the incrementality property) has only been discussed by Holmes and Forster (1970) in this context. The originators of this technique, Abrams and Bever (1969), explained their data solely in terms of what they called the end-of-clause effect (see, also, Bever and Hurtig, 1975), but the two factors we have analyzed here seem to be clearly operative there too, and that muddies their data significantly. That is, even though these scholars placed a tone at the end of a clause, this tone position constituted the first of a decreasing tendency in a series of three tones, and thus the higher RTs to this (first) position may not have been the sole result of structural factors. There is, therefore, a very possible confusion and conflation between perceptual/cognitive load factors and structural effects in their data, and this merits a closer look.

The two factors we have identified here—the position effect and processing load—conspire to yield the RTs that can be obtained with the tone monitoring technique, and as a result future experiments employing this technique, we advise, will need to take this contingency into consideration. In our study we have shown that the two factors can be certainly separated, especially when one sets out to do so, but these factors may hide or obscure structural effects in tone monitoring tasks, requiring a more focused methodological design if structural effects constitute the focus point. The interaction between processing load and the position effect is also discernible in previous studies, even if it seems to not have been explicitly noticed, and structural effects surface only in specific circumstances, as we shall now show.

Take the experiments in Cohen and Mehler (1996) as a case in point, which we now approach through an analysis that speaks to our concerns; in particular, we will show that

in this study too differences in RTs disappear across differing structural conditions when the position effect overcomes these experimental conditions, the unnoticed factor we are discussing. It is hoped that this analysis further illuminates the main point of our study, and put an end to this paper to boot.

In the first three experiments reported in Cohen and Mehler (1996), length was controlled for across two types of sentences and different RTs were recorded in the same position across these sentence types, which naturally suggests purely structural effects. Cohen and Mehler (1996), however, used relative clauses, which are certainly more complex than the non-relative structures that both Abrams and Bever (1969) and Holmes and Forster (1970) employed. That the position effect appears to have been nullified in Cohen and Mehler (1996) may be the result of pushing memory resources to the limit, perhaps the key to eliciting structural effects in the click-detection task. A closer look at their materials will further illustrate.

In the first experiment, Cohen and Mehler (1996) compared reversible subject and transposed object relatives in French, a pertinent pair to compare given that in this condition the complementizer is the only differing element between the two sentences (*qui* in subject relatives, *que* in the object constructions). Consider the following pair, where the | symbol marks where the tone was placed and the numbers within brackets indicate the RTs (the translation for these French sentences can be found in the original paper).

- (3) Le savant (qui connaît le docteur) t|ravaille... (218 ms.)
- (4) Le savant (que connaît le docteur) t|ravaille... (234)

Note that the RTs to a tone placed right after the embedded clause indicates, in agreement with the literature (see some of the references mentioned in Gibson, 1998), that object relatives are harder to process than subject relatives. In a second experiment, these results were replicated (the RTs were 248 and 272, respectively) and then compared to a new tone position: right before the end of the embedded clause.

- (5) Le savant (qui connaît le d|octeur) travaille... (249 ms.)
- (6) Le savant (que connaît le d|octeur) travaille... (250)

Interestingly, RTs to tones before the end of a clause are not different across sentence type, and perhaps this is not unexpected, given that the adjustment processing a relative clause comports supposedly takes place once the internal phrase has been completed, and not before. If so, it seems that the greater cognitive load an object relative exerts is in fact operative *after* the embedded clause has been processed, but not during it. In the third experiment, though,

the object relative was presented in its natural canonical order (i.e., it was not transposed), tones were placed after the embedded clauses, and the differences in RTs disappeared altogether:

- (7) Le savant (qui connaît le docteur) t|raivaille... (262 ms.)
- (8) Le savant (que le docteur connaît) t|raivaille... (264)

The last datum is interesting, as it suggests that object relatives in their normal manifestation are not more difficult to process than subject relatives—at least while participants monitor a tone.⁷ We submit that when across-sentence-type differences disappear, this is (mostly) due to the position effect, suggesting that in some cases structural differences and the placement of tone positions—the factors Cohen & Mehler manipulated—are not as pronounced as to mitigate the strength of perceptual factors.

As we argued, this seems to have been the case in Experiments 1a and 1b, where the structural differences between the two types of sentences we employed were minimal and the tone positions did not correlate with operations that might have disrupted the decreasing tendency of RTs (as was the case in Experiment 3). If so, the position effect has a great role to play in the explanation of the response patterns in tone-monitoring tasks in general, but this has so far been missed. More precisely, the design of a monitoring experiment appears to influence whether the position effect is operative or not—and to what extent it is.

The take-home message, then, is that reacting to a tone demands non-trivial attentional resources. We have shown that, on the one hand, two very different factors have a great influence upon response patterns in general; and on the other hand, that designing a tone-monitoring experiment requires careful consideration if structural effects are being sought—the right balance needs to be struck between the quantity and complexity of the experimental materials and the memory threshold that is involved in processing them at specific locations. As things stand, the results of Abrams and Bever (1969) and Bever and Hurtig (1975) in particular do not appear to us to be on very solid ground any more—is there

⁷A caveat. We do not have any data for the last possible combination: tones inside the embedded clause in the normal, canonical order. It could well be the case that the greater processing load of objective relatives is present after the embedded clause in the transposed condition but before it in the canonical order. Somewhat relevantly, Mota and Igoa (2014) report a tone monitoring experiment with two types of sentences—embedded clauses with either multiple prepositional phrases or multiple conjunctive phrases—in which RTs to a series of tones inside the internal phrases of the embedded clause increase instead of decrease. This is entirely in line with the claim that RTs only decrease if linguistic uncertainty decreases too, as this is certainly not the case the deeper into an embedded clause the parser goes (and we note that these authors did not find any across-sentence-type differences, either). We expect that a similar pattern would emerge in the combination Cohen & Mehler did not run.

really an end-of-clause effect in those data or just a position effect?—and a revisit to biclausal sentences and the like seems in order. We hope to contribute to that line of work in the near future, and we certainly expect the present study to be of much relevance to the sort of issues Franco et al. (2015) addressed.

Acknowledgements The research presented here was funded, at least in part, by a Beatriu de Pinós fellowship awarded by the Catalan Research Council (AGAUR) to the first author (2011-BP-A-00127), an AGAUR research grant awarded to the Psycholinguistics Research Group at the Universitat Rovira i Virgili in Tarragona, Spain (2014-SGR-1444), and a grant from the Spanish Ministry of Economy and Competitiveness (PSI2015-63525-P).

References

- Abrams, K., & Bever, T. G. (1969). Syntactic structure modifies attention during speech perception and recognition. *The Quarterly Journal of Experimental Psychology*, 21(3), 280–290.
- Almela, R., Cantos, P., Sánchez, A., Sarmiento, R., & Almela, M. (2005). *Frecuencias del Español. Diccionario y estudios léxicos y morfológicos*. Madrid: Editorial Universitas.
- Bever, T. G. (1973). Serial position and response biases do not account for the effect of syntactic structure on the location of brief noises during sentences. *Journal of Psycholinguistic Research*, 2, 287–288.
- Bever, T. G., & Hurtig, R. R. (1975). Detection of a nonlinguistic stimulus is poorest at the end of a clause. *Journal of Psycholinguistic Research*, 4(1), 1–7.
- Bever, T. G., Lackner, J. R., & Kirk, R. (1969). The underlying structures of sentences are the primary units of immediate speech processing. *Perception and Psychophysics*, 5(4), 225–234.
- Chapin, P. G., Smith, T. S., & Abrahamson, A. A. (1972). Two factors in perceptual segmentation of speech. *Journal of Verbal Learning and Verbal Behavior*, 11, 164–173.
- Cohen, L., & Mehler, J. (1996). Click monitoring revisited: An on-line study of sentence comprehension. *Memory and Cognition*, 24(1), 94–102.
- Cutler, A., & Norris, D. (1979). Monitoring sentence comprehension. In Cooper, W. E., & Walker, E. C. T. (Eds.) *Sentence processing: Psycholinguistic studies presented to Merrill Garrett* (pp. 113–34). Hillsdale: Lawrence Erlbaum.
- Flores d'Arcais, G. B. (1978). The perception of complex sentences. In Levelt, W. J. M., & Flores d'Arcais, G. B. (Eds.) *Studies in the perception of language* (pp. 155–185). New York: Wiley.
- Fodor, J. A. (1983). *The modularity of mind*. Cambridge: Bradford Books/The MIT Press.
- Fodor, J. A., & Bever, T. G. (1965). The psychological reality of linguistic segments. *Journal of Verbal Learning and Verbal Behavior*, 4, 414–420.
- Fodor, J. A., Bever, T. G., & Garrett, M. F. (1974). *The psychology of language*. London: McGraw-Hill.
- Forster, K. I., & Forster, J. C. (2003). DMDX: A windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, and Computers*, 35, 116–124.
- Franco, A., Gaillard, V., Cleeremans, A., & Destrebecqz, A. (2015). Assessing segmentation processes by click detection: Online measure of statistical learning, or simple interference? *Behavior Research Methods*, 47(4), 1393–1403.
- Garrett, M. F., Bever, T., & Fodor, J. A. (1966). Active use of grammar in speech perception. *Perception and Psychophysics*, 1, 30–2.
- Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition*, 68, 1–76.
- Giraudet, L., St-Louis, M.-E., Scannella, S., & Causse, M. (2015). P300 event-related potential as an indicator of inattention deafness? *PLoS ONE*, 10(2), 1–18.
- Gómez, D. M., Bion, R. A. H., & Mehler, J. (2011). The word segmentation process as revealed by click detection. *Language and Cognitive Processes*, 26(2), 212–223.
- Greenhouse, S. W., & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika*, 24, 95–112.
- Harley, T. (2001). *The psychology of language*. Sussex: Psychology Press.
- Holmes, V. M., & Forster, K. I. (1970). Detection of extraneous signals during sentence recognition. *Perception and Psychophysics*, 7(5), 297–301.
- Just, M. A., Carpenter, P. A., & Woolley, J. D. (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology*, 111(2), 228–238.
- Käthner, I., Wriessnegger, S. C., Müller-Putz, G. R., & Kübler, A. (2014). Effects of mental workload and fatigue on the p300, alpha and theta band power during operation of an ERP (p300) brain-computer interface. *Biological Psychology*, 102, 118–129.
- Kok, A. (2001). On the utility of p3 amplitude as a measure of processing capacity. *Psychophysiology*, 38, 557–577.
- Levelt, W. J. M. (1978). A survey of studies in sentence perception: 1970–1976. In Levelt, W. J. M., & Flores d'Arcais, G. B. (Eds.) *Studies in the perception of language* (pp. 1–74). New York: Wiley.
- Moro, A. (2008). *The boundaries of babel*. Cambridge: The MIT Press.
- Mota, S., & Igoa, J. M. (2014). Recursion in syntactic processing in Spanish. In *IEICE Technical Report* (pp. 1–6).
- Näätänen, R., & Picton, T. (1987). The n1 wave of the human electric and magnetic response to sound: A review and an analysis of the component structure. *Psychophysiology*, 24(4), 375–425.
- Pickering, M. J., & van Gompel, R. P. G. (2006). Syntactic parsing. In Traxler, M., & Gernsbacher, M. A. (Eds.) *The handbook of psycholinguistics* (pp. 455–503). London: Academic Press Inc.
- Reber, A. S., & Anderson, J. R. (1970). The perception of clicks in linguistic and nonlinguistic messages. *Perception and Psychophysics*, 8(2), 81–89.
- Sebastián-Gallés, N., Martí, M. A., Carreiras, M., & Cuetos, F. (2000). *LEXESP. Léxico informatizado del español*. Barcelona: Edicions Universitat de Barcelona.
- Sirevaag, E. J., Kramer, A. F., Wickens, C. D., Reisweber, M., Strayer, D. L., & Grenell, J. F. (1993). Assessment of pilot performance and mental workload in rotary wing aircraft. *Ergonomics*, 36(9), 1121–1140.
- van Gompel, R. P. G., & Pickering, M. J. (2009). Syntactic parsing. In Gaskell, G. (Ed.) *The Oxford handbook of psycholinguistics* (pp. 289–307). Oxford: Oxford University Press.
- Wickens, C., Kramer, A., Vanasse, L., & Donchin, E. (1983). Performance of concurrent tasks: A psychophysiological analysis of the reciprocity of information-processing resources. *Science*, 221(4615), 1080–1082.