Prosody, Performance, and Cognitive Skill: Evidence from Individual Differences

Fernanda Ferreira and Hossein Karimi

Abstract If a pause occurs in the middle of a sentence, is it attributable to prosodic structure, planning problems, or both? And if both prosodic representation and performance constraints conspire to cause a speaker to divide a sentence into two units, can the durational effects that result be parsed into those two different sources? In this chapter, we argue that prosody and performance are theoretically and empirically distinct, and that durational effects may arise from two distinct sources: from the implementation of a grammatical representation, and from performance limitations. A range of empirical evidence is presented to support this distinction. Studies investigating the effects of working memory, inhibitory control, and lexical difficulty indicate that individuals with less cognitive capacity are more likely to produce sentence-internal breaks, and these are not conditioned by characteristics of a prosodic representation. This finding suggests that performance units are not necessarily prosodic units, and that an adequate theory of sentence production must incorporate mechanisms for implementing prosodic structure as well as strategies for managing processing load during speech.

Keywords Language production · Prosody · Timing · Working memory · Inhibition

1 Introduction

When speakers pause in the middle of a sentence, is the pause attributable to the speaker's implementation of a prosodic representation, or do speakers pause for some performance reason—for example, to buy more time to plan the upcoming stretch of speech? Or, is the correct answer "both"? That is, speakers sometimes not only need time to plan or in some way manage their cognitive resources but they

F. Ferreira (✉) · H. Karimi
Department of Psychology, Institute for Mind and Brain,
University of South Carolina, Columbia, SC 29201, USA
e-mail: Fernanda@sc.edu

H. Karimi
e-mail: karimi@email.sc.edu

L. Frazier, E. Gibson (eds.), *Explicit and Implicit Prosody in Sentence Processing,*
Studies in Theoretical Psycholinguistics 46, DOI 10.1007/978-3-319-12961-7_7

also use prosodic information to achieve their performance goals in a linguistically principled way. These are the questions we address in this chapter. Before we begin, however, we would like to highlight the extraordinary influence that Janet Fodor has had on the field of psycholinguistics, not just due to her work on prosody, the focus of the present volume, but also through her contributions to numerous other debates as well. Whether the subject is the online processing of phrase structure (Frazier and Fodor 1978; Fodor and Frazier 1980), the establishment of filler–gap relations (Fodor 1978), the reanalysis of garden-path sentences (Fodor and Inoue 1994, 1998, 2000; Fodor and Ferreira 1998), or the critical role of prosody in written and spoken language (Fodor 2002a, b), Janet Fodor's arguments have sharpened the issues and allowed psycholinguists from a variety of perspectives and theoretical orientations to design coherent and theoretically focused experiments and to draw conclusions that genuinely move the field forward. This is very clearly true when it comes to the role of prosody in language processing. Let us now turn to this topic.

It is probably fair to say that psycholinguists tend to be biased towards what we might term "naturalized prosody"—that is, they are predisposed to believe that prosodic effects arise, at least in part, due to factors related to performance in sentence planning. But in our own work (Ferreira 1991, 1993, 2007), we have adopted the strong position that prosody and performance effects must be distinguished in any psycholinguistic model. In that early work (Ferreira 1991, 1993), we found empirical evidence suggesting that the left and right contexts surrounding a word have markedly different effects on word and pause durations: The complexity of upcoming material influenced the likelihood of a pause but did not lead to word lengthening. In addition, pause durations patterned with sentence initiation times. In contrast, the prosodic complexity of the context to the left of a word affected that word's duration and what we characterized as grammar-based pauses: Pauses of a relatively short duration that tend to co-occur with phrase-final lengthening. We also argued that these pauses arise in part because a syllable reaches the limits of its "stretchability," and as a result, the speaker is unable to maintain a timing pattern with lengthening alone (Ferreira 1993; Selkirk 1984). We therefore concluded that acoustic effects associated with material to the left of a potential prosodic boundary are related to implementation of a metrical representation, whereas those associated with material to the right are attributable to planning and performance factors. This model which assumes prosodic effects from left context and planning effects from right context has been challenged based on new processing models that offer more sophisticated accounts of how performance constraints might lead to prosodic effects (Watson and Gibson 2004). Nonetheless, we have maintained that these newer algorithms and findings are not entirely persuasive, in part because the success of any algorithm depends critically on the choice of sentences used to evaluate it (Ferreira 2007), and most studies do not employ a design in which left and right contexts are systematically and orthogonally manipulated.

In contrast to performance, we view prosody as a linguistic system with its own grammar. The grammar has a metrical component, which causes an utterance to have a distinct and grammatical rhythm, and an intonational component, which is meant to capture changes in pitch across an utterance. Both are a function of prosodic constituency, which is derived from rules that define the syntax–phonology interface. These prosodic domains in turn determine the application of rules linked to phrase-final lengthening and pausing, as well as the placement of pitch accents associated with different communicative intentions. Thus, a set of grammatical constraints defines prosodic structure and the rules that apply to those structures. One interesting aspect of prosody is that the application of rules is often graded, with optimality theory approaches (Prince and Smolensky 1993) well suited to capturing the idea that rules do not apply in an all-or-none manner, but instead apply with a certain probability depending on the precise balance of conflicting constraints. In addition, prosodic constituents are created from a syntactic structure, but the two forms of representation are not isomorphic. Both these points will become relevant when at the end of the chapter we consider the viability of a hybrid approach relating prosody and performance.

On the other hand, performance effects are often poorly behaved with respect to any semantic or syntactic constraints that might govern the production of a sentence. To take a clear example, if a person pauses after a sentence initial *the,* which is fairly common in spontaneous speech (Boomer 1965; Maclay and Osgood 1959), that pause has no obvious grammatical motivation. Indeed, the pause would usually be treated as a disfluency, and a speaker aiming to speak fluently would avoid it. This is not to say that the disfluency is random and conveys no information to the listener. A fair bit of research has shown that listeners in fact can use disfluencies as information concerning what might be coming up next, and these predictions are based on listeners' knowledge of typical co-occurrences between, for example, difficult concepts and the need to pause to allow time for lexical retrieval (Arnold et al. 2007). Nonetheless, few would think of such a pause as prosodically conditioned.

This case seems clear-cut, but the picture gets a little more complicated when we consider pauses in other sentence locations, especially near the middle of an utterance. Consider this example: *Mary ordered salad because <pause> she is trying to eat more healthily*. The pause after *because* is not in the syntactically most prominent location; *because* is part of the second clause, and therefore the pause after it separates *because* from the syntactic constituent of which it is a part. Based on this criterion, it might be tempting to view a pause in this location as also non-prosodic, like the one after a sentence-initial *the*. On the other hand, if we make reference to constraints on prosodic rather than syntactic constituency, then that pause location is perfectly fine. As was argued decades ago for cases such as *This is the cat that chased the rat that swallowed the cheese…* (Chomsky and Halle 1968), the rules of phonology seem to have the effect of simplifying and flattening a syntactic structure. And as argued more recently by Selkirk (1984), a particular intonational phrasing is acceptable as long as the resulting phrases obey the sense unit condition, which states that the constituents inside an intonational phrase must be in a head–modifier or head–argument relation. In addition, the phrasing that groups *because* with the first clause has the additional virtue of dividing the sentence into two parts of roughly equal size—two balanced sisters, to use Fodor's terminology (Fodor 1998, 2002a, b). Thus, in this sort of example, it is more difficult to tell whether the pause is prosodic, performance based, or both. Careful experiments are required to distinguish the two potential sources of lengthening and pausing—the grammar versus disfluency.

In this chapter, we approach the distinction between prosody and performance in a somewhat novel way: We will motivate the distinction by reviewing evidence from new research that links performance effects to cognitive skill. These studies use an individual differences approach to assess whether people with lower working memory (WM) capacity, weaker inhibitory control, or lower intelligence quotients (IQs) are also more likely to need a break point within a sentence compared to those with more robust cognitive systems. The logic of the approach is to assume that there is no principled reason to expect that prosodic effects will be influenced by cognitive skill; the grammar is the grammar whether a person has high or low WM capacity. Of course, in cases in which the grammar presents the language system with a choice between more than one linguistic structure, cognitive factors will play an important role in making the linguistic choice. The grammar presents options, and the cognitive system selects from among them on the basis of a range of factors, including performance constraints. In contrast, performance is clearly affected by cognitive skill. For example, a person who has a shorter WM span would seem to be more likely to break up a sentence into smaller performance units than would someone with a longer span. We turn to these studies next.

2 Working Memory and Implicit Prosody

One of the most influential and important ideas to emerge from psycholinguistics in the past decade or so is the notion that prosody is not confined to spoken language: Readers also generate a prosodic representation for written sentences. This proposal is compatible with decades of research in cognitive psychology showing that, fundamentally, reading is the translation of visual symbols into a phonological code (Berent and Perfetti 1995). Visually presented words activate their phonological forms, as demonstrated by phenomena such as tongue-twister effects in reading (McCutchen and Perfetti 1982) as well as interference from homophones (Van Orden 1987). For instance, using a semantic categorization task, Van Orden demonstrated that homophones associated with a target significantly increased false positive categorization rates. He observed that the word *rows* was sometimes miscategorized as a kind of flower, indicating not only that the phonological representation of words are activated during reading but also that this representation might mediate access to the word's semantic representation. Reading also seems to involve an "inner voice" that generates an ongoing phonological representation of text, and which even has speech characteristics such as gender (Quinn et al. 2000; Slowiaczek and Clifton 1980; Stolterfoht et al. 2007). What Fodor added to our theoretical understanding of phonological processing during reading is critical for psycholinguistics: She explicitly argued that the sounds we hear as we read include prosodic information, and are governed by a principled representation of prosodic structure. Many of these arguments were based on studies of garden-path reanalysis, most of which were conducted with visual materials, and many of which suggested that revision of an incorrect syntactic structure is more difficult when the new analysis requires the generation of a different prosodic form for the sentence. In one seminal study using self-paced reading, Bader (1998) used focus operators to manipulate the prosodic structure of local garden-path ambiguities and showed that prosodic structure can influence recovery from a misanalysis independent of syntactic structure, suggesting that reanalysis is prosodically constrained, and more importantly for our purposes, providing evidence for implicit prosody in reading.

Armed with this theoretical innovation, we conducted a large-scale individual differences study designed to investigate the relationship between WM capacity and attachment decisions (Swets et al. 2007). Our research strategy was to identify a long sentence type that would likely need to be spoken as more than one production unit. For this purpose we chose the relative clause attachment structure illustrated in *The maid of the princess who scratched herself in public was terribly embarrassed*. This sentence seems to allow for two possible break points: one after *public,* at the subject–verb phrase boundary, and the other after *princess,* before the relative clause. These options are rank ordered, of course: The location between the subject and verb phrase is the one that is structurally most preferred, and the location before the relative clause might also be exploited if an individual has such limited processing capacity that he/she must divide the sentence into more than two performance units. In addition, as has been widely discussed, the sentence is globally ambiguous because the relative clause can attach either high, to the first noun (N1, *maid*), or low, to the second and more recent noun (N2, *princess*). The preference for N1 or N2 attachment seems to vary crosslinguistically: Dutch has about a 60–40 % bias for N1 attachments (Desmet et al. 2002), whereas English has a 40–60 % bias for N2 attachments (Cuetos and Mitchell 1988). These crosslinguistic differences have been explained by appealing to implicit prosody: Whereas speakers of Dutch tend to put a prosodic break between the complex noun phrase (NP) and the relative clause in sentences like these, English speakers tend to leave out this break and prefer to place a break after the relative clause instead of before it. If a speaker does insert a prosodic break before the relative clause, as Dutch speakers tend to do, the result is a bias towards higher attachment decisions for spoken sentences (Carlson et al. 2001). The prosodic break is assumed to induce N1 preferences because it can be interpreted as a "structural discontinuity in the syntactic tree" (Fodor 2002a, p. 4). This interpretation supports the formation of a tree in which the entire NP is modified by the relative clause rather than just N2, resulting in a high-attachment (N1) preference.

Speculation has also centered around whether the preference for N1 versus N2 attachment might be related to WM capacity. The intuitive idea is that recency favors N2 attachment, and those with smaller working memories might be more biased to use a recency strategy to make attachment decisions, as reported by Felser et al. (2003) for 6–7-year-olds. This possibility has even occasionally been invoked to explain the crosslinguistic differences in attachment preference mentioned above, with the argument going something like this: There is a tendency for the N1 preference to be found in experiments which include participants attending European universities, and for the N2 preference to emerge in experiments with students from American universities, especially large public institutions. If we assume that selectivity is correlated with WM capacity (and there is evidence that WM and IQ are positively correlated), then perhaps what appears to be a crosslinguistic difference is actually a confound caused by testing participant groups with different individual difference characteristics. This would be an unfortunate interpretation, but fortunately, the results of Swets et al. (2007) allow us to rule it out, as we will see shortly.

The study was unusual (perhaps unique at the time) for adopting a psychometric approach to these psycholinguistic questions concerning attachment preference and implicit prosody. A psychometric approach attempts to establish relationships among variables that occur naturally and that naturally vary (i.e., WM capacity), in contrast with variables that can be experimentally manipulated. The statistical method is then to test for correlations using sophisticated quantitative techniques such as structural equation modeling. An important requirement of such work is that sample sizes be adequate to ensure there is sufficient power to conduct continuous analyses because continuous analyses allow researchers to evaluate the relationship between two variables across the full range of scores and allow them to avoid the problems inherent in the use of so-called extreme-groups designs (i.e., the testing of only the subjects with the highest and the lowest WM scores, so that WM is treated as a categorical variable in statistical analyses). To that end, 150 Michigan State University undergraduates, all native speakers of English, were tested along with 96 undergraduates from Ghent University, all of whom were native speakers of Dutch. Each person's WM capacity was assessed using a reading span task and a separate spatial span task. Then participants were shown sentences individually, and after each sentence, the participants answered a question such as *Who scratched herself in public,* with the options represented by N1 and N2 attachments shown one above the other.

The critical manipulation in this study was conducted between experiments as well as between subjects. In Experiment 1, each sentence was presented on a single line, so that nothing about the visual presentation encouraged the inclusion of a prosodic break within the sentence. With this setup, Swets et al. (2007) replicated previous work showing that Dutch participants prefer to attach to N1 and English participants to N2; however, although the effect was statistically significant, it was quite a bit smaller than in previous studies, amounting to no more than a 3–4% difference in attachment decisions. Much more surprising was the effect of WM: Contrary to the recency principle, we observed that the smaller a participant's WM capacity, the more likely he or she was to prefer N1 attachments. Moreover, this effect of WM was statistically identical for Dutch and English participants, suggesting that it was entirely independent of any crosslinguistic factors. Moreover, if the participants are divided into two equal *n* groups based on their WM capacities, the pattern that emerges is that the participants with the lowest spans preferred N1 attachments whether they were English or Dutch, and the participants with the highest spans preferred N2 attachments, again regardless of what language they spoke.

How do we explain this counterintuitive result? Our account made a critical appeal to the notion of prosodic chunking in silent reading. Imagine that high-span subjects can "chunk" more information together while reading. These higher-span individuals are able to treat the entire subject of the sentence as a single "processing unit". Low-span readers, in contrast, may have to break up the subject because of its length. A likely boundary for such a break point is right before the relative clause. And these breaks in turn will encourage N1 attachments, for the reasons described earlier. This hypothesis that chunking strategies underlie the individual differences observed in our first experiment was tested in the second experiment. This time, the sentences were presented in three successive displays. The first included the words before the relative clause (*The maid of the princess*), the second consisted of the entire relative clause (*who scratched herself in public*), and the third consisted of the entire verb phrase (*was terribly embarrassed*). Our prediction was that this segmented presentation method would induce readers to prosodically chunk the sentences into three units, including one that separated both potential attachment sites from the relative clause. As a result, all participants would be turned into low-capacity readers; based on the presentation format, all participants would generate an implicit prosodic phrasing that isolated the relative clause, and based on the principles mentioned earlier, this would lead to an overall preference for N1 attachments.

These predictions were clearly confirmed. Although we once again replicated the slight preference for N1 attachments in Dutch and for N2 attachments in English, we no longer observed a significant effect of WM capacity in either group. Not only did everyone regardless of WM capacity prefer N1 attachments but also the overall N1 preference was much larger than has been reported in any previous work: 71 % for English speakers and 75 % for Dutch speakers (the two groups did not differ significantly). Thus, if we manipulate presentation format so that all participants are induced to read the way we hypothesize low-capacity readers do, we dramatically enhance the N1 attachment preference.

Two important conclusions can be drawn from this study. First, we have discovered some of the strongest evidence we know of for the reality of implicit prosody in reading. Moreover, in pilot work we are currently conducting in our laboratory, we are measuring WM capacity once again, but this time asking participants to say the sentences out loud. Participants are being asked to read and learn the sentences, and then to repeat them from memory upon receipt of a cue. We additionally varied whether the verb in the relative clause was high or low frequency (e.g., *glorified* vs. *idolized*), because we predicted that greater lexical difficulty would increase the chances of a performance break, particularly before that relative clause. Our preliminary data suggest that people with lower spans are more likely to require two break points within these same sentences, and that they are also more affected by the frequency manipulation. It thus appears that our findings concerning implicit and explicit prosody dovetail nicely: Regardless of whether people speak out loud or read silently, it seems that those with smaller WM spans are more likely to divide a sentence up into multiple performance units. This is our first important conclusion. Second, given this relationship between WM capacity, which is a cognitive ability factor, and the tendency to break up a sentence, we think it makes a great deal of sense to think of these units not as prosodic constituents but as performance units. We base this argument on the idea that prosodic constituency has no obvious connection to cognitive capacity; there is no theoretical reason for believing that WM span is in any way related to the way the grammar of prosody is applied or implemented. In contrast, there are very compelling theoretical reasons for linking WM and performance; indeed, in multiple domains it has been observed that those with larger spans chunk information more effectively and are able to pack more information into a single chunk (Ottem et al. 2007).

In short, the chunks formed during silent reading are affected by WM capacity, as would be expected if performance units reflect cognitive skill. This in turn motivates a separation between prosodic and performance-based effects in language processing.

3 Inhibitory Control and Planning in Production

Next, we turn to research we have conducted investigating the relationship between the integrity of inhibitory systems and speakers' fluency. Broadly speaking, inhibition as a cognitive skill can be defined as the suppression of inappropriate responses or intervening memories when the context changes (Aron et al. 2004). In other words, cognitive inhibition is a mechanism whereby prepotent behavioral responses are constrained when the expression of such responses is inappropriate or incorrect (Burle et al. 2004). A powerful method for investigating inhibition is again to use an individual differences approach—in this case, to compare performance in individuals suffering from attention deficit hyperactivity disorder (ADHD) to the performance of demographically matched controls (people of approximately the same age and social/educational background). A large number of studies suggest that people with ADHD have impaired inhibitory systems, leading to problems in tasks such as the anti-saccade and Stroop task, both of which require participants to squelch a prepotent response. For example, in the anti-saccade task, participants are instructed to look away from a visual stimulus such as a cross or dot as soon as it appears on the screen (Hallet 1978; Nieuwenhuis et al. 2001). Because such an abrupt onset of visual stimulus is known to automatically capture attention and eye movements (Theeuwes et al. 1998), efficient anti-saccade performance requires inhibition of the reflexive eye movement towards the abrupt-onset stimulus (Nieuwenhuis et al. 2001). Our work with this population has also shown that ADHD is characterized by more focused inhibitory deficits related specifically to language planning. In one study (Engelhardt et al. 2010), we asked individuals with ADHD as well as matched control subjects to generate a sentence from two objects (one animate, one inanimate), together with a printed verb. The verb either was ambiguous between simple past and past participle (*moved*) or was unambiguously the past participle (*ridden*) form, and presentation of the animate object (e.g., the girl) either preceded or followed the presentation of the inanimate object (e.g., the bicycle). Thus, in the past participle condition where the animate entity was presented first, the participants could start uttering the sentence "the girl…" and then realize at this point that the sentence needs to be in passive form ("The bicycle was ridden by the girl."). As predicted, we observed that both groups of subjects were less fluent producing sentences with the participle verb, particularly when an animate object was shown before the inanimate object. This is because the participle nearly forces the generation of a passive form (the past perfect is a legitimate alternative, but our participants seemed to be unaware of this), which in turn forces the inanimate entity to serve as the sentential subject. In addition, as would be expected, given that people with ADHD tend to have problems with inhibitory control, this effect was larger for those with ADHD. The effect was particularly pronounced for self-repairs, suggesting that problems with inhibition lead individuals with ADHD to begin speaking before they have planned out the entire utterance and know it will be grammatical and semantically appropriate. Post hoc analyses also revealed that lower IQ scores were associated with more disfluencies overall, perhaps because one component of the IQ is vocabulary knowledge, which presumably relates to the ease of retrieving information from the lexicon.

In follow-up research, Engelhardt et al. (2011) asked healthy subjects and matched individuals with ADHD to describe networks of colored circles so that another person could draw the networks based on those descriptions. The resulting utterances had this character: *First there is an orange dot, and above it is a red dot. To the left of the red dot is a green dot and a blue dot,* etc. Successful description of these networks required some planning because the networks contained branches and choice points, and therefore speakers had to decide which branch of the network to describe next, and they had to make sure they remembered the choice circle so they could return to it to describe its other branches. In contrast to the previous study, this one taps into sentence planning at a level higher than grammatical encoding. Based on our other work, however, we expected to find that people with ADHD generated less fluent descriptions, and this is what we reported in the study: People with ADHD paused more often and generated more self-repairs than did normal controls. These differences were observed even though the two groups were matched on age, IQ, years of education, and even reading ability.

Thus, it appears that weaker inhibitory systems are associated with more errors and pauses in language production. We will make the same argument concerning inhibition that we made earlier with respect to WM: There is no theory of prosody from which predictions concerning effects of inhibition fall out naturally. Again, prosody is part of the grammar, and the grammar does not appeal to factors relating to cognitive skill. In contrast, there are compelling reasons for thinking that cognitive skill—in this case, inhibitory control—would be associated with performance and the need to pause or break during language production. This leads us to conclude that prosody and performance are distinct phenomena: Prosody is about the grammar, whereas performance is influenced by individual difference characteristics relating to cognitive skill.

Finally, in a recent study of individual differences among 106 normal participants, we used structural equation modeling to assess the relationships between various cognitive skills and the tendency to be disfluent during production (Engelhardt et al. 2013). This study included a range of measures of both intelligence (e.g., processing speed, vocabulary) and executive control (e.g., a stop-signal reaction time task and a Stroop task). We observed no significant effects related to IQ once correlated relationships with executive control were statistically removed, but we found a moderate effect of executive functioning, suggesting that those with poorer executive control and, in particular, those with weaker inhibitory systems tended to be less fluent. Thus, it is not only in clinical populations that we find a relationship between cognitive skill and fluency but we also see that within a large group of normal speakers, those with less intact cognitive systems are more likely to have performance problems during production.

In summary, then, factors that are not naturally thought of as part of the grammar seem to have a strong effect on language performance. We have seen that both smaller WM capacity and weaker inhibitory control systems cause speakers to produce more pauses and breaks when they speak. From these data, we argue that prosody and performance are distinct phenomena, and therefore no adequate theory of language production or of prosody in psycholinguistics can conflate them—to do so would be to blur important distinctions among representational types and processing mechanisms.

4 Bringing Prosody and Performance Together

Having laid out our arguments for distinguishing prosody and performance, we now want to consider how we can think about the interactions between the two, and the way both affect the auditory characteristics of a sentence. As we argued previously (Ferreira 1993, 2007), if we measure a variable such as pause duration, any effects are likely to be a mixture of both planning and rhythm—some of the pause time is attributable to the need to plan upcoming material, and some of it is attributable to the implementation of a prosodic representation and the need to insert pauses in order to maintain a specified rhythm.

Planning-based pauses are typically longer than prosodic pauses, and also tend to correlate with other planning-based variables such as sentence initiation time. In addition, these pauses will tend to get shorter and eventually disappear as a speaker becomes more practiced and fluent with a particular utterance. In contrast, rhythmic pauses are shorter, correlated with other prosodic effects such as phrase-final lengthening, and, by hypothesis, cannot be deleted without harming the prosodic well-formedness of the utterance. One way to think about the distinction is with an analogy to music: When a musician plays a piece of music, she will insert pauses at particular locations as she attempts to implement the musical score, and of course rests in specific places and of specific durations are as integral to any musical piece as the notes are. But if she struggles a bit with a certain sequence of notes and pauses before trying to execute them in order to plan the movements, that pause is a performance-based pause and ultimately needs to be smoothed out if the musician wants to give a performance that will be viewed as competent and aesthetically pleasing. Rests, then, remain in the performance, but silences due to cognitive challenges are essentially errors and need to be eliminated if the musician wants to be viewed as a skilled musician. The same idea applies to language production: Rhythmic pauses must be maintained in an utterance, but planning-based pauses are performance effects and should not be included in the most fluent renditions of the utterance.

At the same time, spontaneous speech will almost invariably be a mixture of both planning and rhythmic effects. Where we want to end this chapter is with a theoretical speculation: Speakers do indeed sometimes need time to plan or in some way manage their cognitive resources when they produce spontaneous utterances, but they also have the ability to use a prosodic representation to achieve their performance goals in a linguistically principled way. As mentioned previously in our description of the prosodic system, the grammar defines a set of prosodic constituents and rules that apply to the resulting representation. One of the aspects of prosody that makes it attractive to psycholinguists is that both constituent structure and rule application tend to be graded, as are almost all phenomena related to human cognition. For example, the division of a sentence into intonational phrases involves both obligatory and optional constraints. The border between a subordinate and a main clause in a sentence must be marked by an intonational phrase boundary, but when it comes to the division between subject and verb phrase, the speaker has the option to place an intonational phrase boundary there or not. Most often we think of the decision to place the boundary as pragmatically conditioned; speakers use intonational phrase boundaries to convey their communicative intentions, including features such as focus, backgrounding, and mood. But the decision may also sometimes be driven by performance considerations: If a sentence is long and the speaker needs to divide it up to say it easily, then he/she might exercise the option of placing an intonational phrase break at the subject–verb phrase boundary. This break would enable him to recover from executing the subject and would also provide time for planning of the rest of the sentence, while at the same time perhaps conveying information to the listener related to the difficulty of the utterance. Thus, the grammar would be available to define an ideal break point from the perspective of prosodic constituency, and performance factors would help to determine whether the option was actually taken.

In addition, because of the nature of the interface between syntax and prosody, the two representational forms are not necessarily isomorphic. One important difference is that prosody (specifically, the sense unit condition) may allow the subject and verb to occur as part of one prosodic constituent and the postverbal constituents to make up another. An example might be a sentence such as *The noisy students left/after we ran out of beer,* which could naturally be spoken in such a way that the prepositional phrase constitutes its own prosodic phrase. This freedom to deviate from syntactic constituency means that the prosodic system presents the speaker with another tool for managing cognitive load during production: If the subject is relatively short and postverbal material is long, the speaker can create an utterance with two balanced sisters by exercising the option to break after the verb instead of before it. This would result in a more prosodically appealing rendition of a sentence, because sisters that are mismatched in length sound a bit odd, and it would also permit a more even distribution of information over a sentence, an idea consistent with the so-called uniform information density (UID) hypothesis (Jaeger 2010), which assumes that speakers try to avoid major peaks and troughs of information in their utterances, and instead attempt to distribute information more evenly.

Yet another situation that may arise and that is more complex than the others is one that highlights the potential dependencies among different break locations. Let us consider again the relative clause attachment sentences that we focused on earlier; e.g., *The maid of the princess ^ who scratched herself in public ^ was terribly embarrassed,* with the ^ symbol indicating the two potential sites for a prosodic boundary. As we saw previously, a speaker with more limited WM resources might not be able to handle the entire subject as one prosodic unit, and might therefore place a break before the relative clause. But notice that if a speaker chooses this particular site for a prosodic boundary, he has also committed himself to placing a boundary at the subject–verb phrase location as well. This is because (*The maid of the princess, who scratched herself in public was embarrassed*) is not a well-formed prosodic phrasing; it violates rules of prosodic constituency, perhaps creating a "prosodic monster" (Féry, this volume). Thus, the speaker might choose the earlier boundary for reasons related to constraints on cognitive processing, but he/she might then choose the later boundary to maintain prosodic integrity. The first break site would thus be planning based, and the other would be prosodically motivated and even forced. Perhaps these two sources for the two breaks would cause the boundaries to have different properties relating to pitch and other prosodic features, although any differences would be hard to distinguish from those associated with the break locations within the prosodic constituency. Thus, we can argue that the role of the grammar is to create a prosodic representation that gives the cognitive system options when it needs to select.

References

Arnold, J. E., Hudson Kam, C. L., & Tanenhaus, M. K. (2007). If you say thee uh—you're describing something hard: The on-line attribution of disfluency during reference comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33,* 914–930.

Aron, A. R., Robbins, T. W., & Poldrack, R. A. (2004). Inhibition and the right inferior frontal cortex. *Trends in Cognitive Sciences, 8,* 170–177.

Bader, M. (1998). Prosodic influences on reading syntactically ambiguous sentences. In J. D. Fodor, & F. Ferreira (Eds.), *Reanalysis in Sentence Processing* (pp. 1–46). Dordrecht: Kluwer Academic.

Berent, I., & Perfetti, C. A. (1995). A rose is a reez: The two cycles model of phonology assembly in reading English. *Psychological Review, 102,* 146–184.

Boomer, D. S. (1965). Hesitation and grammatical encoding. *Language and Speech, 8,* 148–158.

Burle, B., Vidal, F., Tandonnet, C., & Hasbroucq, T. (2004). Physiological evidence for response inhibition in choice reaction time tasks. *Brain and Cognition, 56,* 153–164.

Carlson, K., Clifton, C., Jr., & Frazier, L. (2001). Prosodic boundaries in adjunct attachment. *Journal of Memory and Language, 45,* 58–81.

Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. New York: Harper & Row.

Cuetos, F., & Mitchell, D. (1988). Cross-linguistic differences in parsing: Restrictions on the use of the late closure strategy in Spanish. *Cognition, 30,* 73–105.

Desmet, T., Brysbaert, M., & Da Baecke, C. (2002). The correspondence between sentence production and corpus frequencies in modifier attachment. *The Quarterly Journal of Experimental Psychology: Section A, 55*(3), 879–896.

Engelhardt, P. E., Corley, M., Nigg, J. T., & Ferreira, F. (2010). The role of inhibition in the production of disfluencies. *Memory & Cognition, 38*(5), 617–628.

Engelhardt, P. E., Ferreira, F., Nigg, J. T. (2011). Language production strategies and disfluencies in multi-clause network descriptions: A study of adult Attention/hyperactivity disorder. *Neuropsychology, 25*(4), 442–453.

Engelhardt, P. E., Nigg, J. T., & Ferreira, F. (2013). Is the disfluency of language outputs related to individual differences in intelligence and executive function? *Acta Psychologica, 144,* 424–432.

Felser, C., Marinis, T., & Clahsen, H. (2003). Children's processing of ambiguous sentences: A study of relative clause attachment. *Language Acquisition, 11,* 127–163.

Ferreira, F. (1991). Effects of length and syntactic complexity on initiation times for prepared utterances. *Journal of Memory and Language, 30,* 210–233.

Ferreira, F. (1993). The creation of prosody during sentence processing. *Psychological Review, 100,* 233–253.

Ferreira, F. (2007). Prosody and performance in language production. *Language and Cognitive Processes, 22*(8), 1151–1177.

Fodor, J. D. (1978). Parsing strategies and constraints on transformations. *Linguistic Inquiry, 9,* 427–473.

Fodor, J. D. (1998). Learning to parse?. *Journal of Psycholinguistic Research, 27,* 285–319.

Fodor, J. D. (2002a). Prosodic disambiguation in silent reading. In: M. Hirotani (Ed.), *NELS 32* (pp. 113–132). Amherst: GLSA Publications.

Fodor, J. D. (2002b). Psycholinguistics cannot escape prosody. *In Proceedings of the SPEECH PROSODY 2002 Conference*, Aix-en-Provence, France, April 2002.

Fodor, J. D., & Ferreira, F. (Eds.). (1998). *Reanalysis in sentence processing*. Dordrecht: Kluwer Academic Publishers.

Fodor, J. D., & Frazier, L. (1980). Is the human sentence processing mechanism an ATN?. *Cognition, 8,* 417–459.

Fodor, J. D., & Inoue, A. (1994). The diagnosis and cure of garden paths. *Journal of Psycholinguistic Research, 23*(4)*,* 405–432.

Fodor, J. D., & Inoue, A. (1998). Attach Anyway. In J. D. Fodor & F. Ferreira (Eds.), *Reanalysis in Sentence Processing*. Dordrecht: Kluwer.

Fodor, J. D., & Inoue, A. (2000). Garden path re-analysis: Attach (anyway) and revision as last resort? In M. D. Vincenzi & V. Lombardo (Eds.), *Cross-linguistic perspective on language processing*. Dordrecht: Kluwer.

Frazier, L., & Fodor, J. D. (1978). The sausage machine: A new two-stage parsing model. *Cognition, 6,* 291–325.

Hallet, P. E. (1978). Primary and secondary saccades to goals defined by instructions. *Vision Research, 18,* 1279–1296.

Jaeger, T. F. (2010). Redundancy and reduction: Speakers manage syntactic information density. *Cognitive Psychology, 61,* 23–62.

Maclay, H., & Osgood, C. E. (1959). Hesitation phenomena in spontaneous English speech. *Word, 15,* 19–44.

McCutchen, D., & Perfetti, C. A. (1982). Coherence and connectedness in the development of discourse production*, Text, 2,* 113–119.

Nieuwenhuis, S., Ridderinkhof, R. K., Blom, J., Band, G. P. H., & Kok, A. (2001). Error-related brain potentials are differentially related to awareness of response errors: Evidence from an antisaccade task. *Psychophysiology, 38,* 752–760.

Ottem, E. J., Lian, A., & Karlsen, P. J. (2007). Reasons for the growth of traditional memory span across age. *European Journal of Cognitive Psychology, 19,* 233–270.

Prince, A., & Smolensky, P. (1993). *Optimality Theory: Constraint Interaction in Generative Grammar*, ms., Rutgers University, New Brunswick, and University of Colorado, Boulder.

Quinn, D., Abdelghany, H., & Fodor, J. D. (2000). *More evidence of implicit prosody in reading: French and Arabic relative clauses*. Poster presented at the 13th Annual CUNY Conference on Human Sentence Processing, La Jolla, March 30–April 1.

Selkirk, E. O. (1984). *Phonology and syntax: The relation between sound and structure*. Cambridge: MIT Press.

Slowiaczek M. L., & Clifton C. (1980). Subvocalization and reading for meaning. *Journal of verbal learning and verbal behavior, 19*(5), 573–582.

Stolterfoht, B., Friederici, A. D., Alter, K., & Steube, A. (2007). Processing focus structure and implicit prosody during silent reading: Differential ERP effects. *Cognition, 104*(3), 565–590.

Swets, B., Desmet, T., Hambrick, D. Z., & Ferreira, F. (2007). The role of working memory in syntactic ambiguity resolution: A psychometric approach. *Journal of Experimental Psychology: General, 136,* 64–81.

Theeuwes, J., Kramer, A. F., Hahn, S., & Irwin, D. E. (1998). Our eyes do not always go where we want them to go: Capture of the eyes by new objects. *Psychological Science, 9,* 379–385.

Van Orden GC. (1987). A rows is a rose: Spelling, sound, and reading. *Memory & Cognition, 15,* 181–198.

Watson, D., & Gibson, E. (2004). The relationship between intonational phrasing and syntactic structure in language production. *Language and Cognitive Processes, 19,* 713–755.