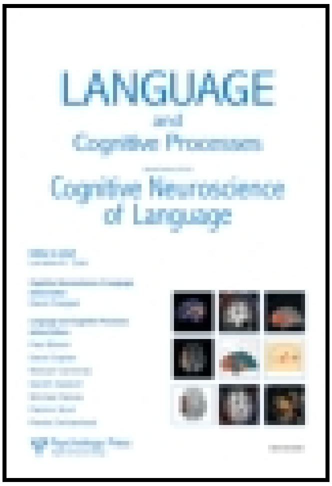

Language and Cognitive Processes
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/plcp20

Prosody and performance in language production

Fernanda Ferreira [a]
[a] School of Philosophy, Psychology, and Language Sciences , University of Edinburgh , Edinburgh, UK

Published online: 17 Dec 2007.

LANGUAGE AND COGNITIVE PROCESSES
2007, 22 (8), 1151–1177

Psychology Press
Taylor & Francis Group

Prosody and performance in language production

Fernanda Ferreira
School of Philosophy, Psychology, and Language Sciences, University of Edinburgh, Edinburgh, UK

Language production theories should explain how speakers generate an utterance's sound structure. One critical question is whether prosody and performance effects have different sources in the production system. It is argued that algorithms designed to predict phenomena such as pauses or intonational breaks are problematic because they tend to conflate prosody and planning. In addition, algorithms that have been proposed have not been evaluated systematically enough to allow their strengths and weaknesses to be assessed and compared, and to the extent that the algorithms have been evaluated, it is clear that they are only moderately successful at predicting the dependent measures of interest. Experimental work suggests that prosodic effects are based on *prosodic constituency* to the left of a potential boundary, and hesitations are due to planning of *syntactic and semantic constituents* to the right. Thus, any adequate algorithm must distinguish between prosody and performance, prosodic and syntactic-semantic constituency, and planning and execution effects.

INTRODUCTION

The aim of this paper is to address whether it is possible to distinguish the prosody of a spoken utterance from acoustic effects that arise as a speaker tries to manage the psychological processes involved in language production. To make this distinction concrete, consider pauses. On the one hand, we know that speakers may pause before beginning a particularly long or difficult utterance (Goldman-Eisler, 1968), just as people often delay initiating any demanding task (Altmann, 2004). On the other hand, it is clear that pauses in speech can be like rests in music, in that they may be used to maintain a rhythmic pattern and thus they may have nothing to do

Correspondence should be addressed to Fernanda, School of Philosophy, Psychology, and Language Sciences, University of Edinburgh, 7 George Square, Edinburgh, EH3 6NN, UK. E-mail: Fernanda.Ferreira@ed.ac.uk

http://www.psypress.com/lcp DOI: 10.1080/01690960701461293

with psychological processes such as planning (Ferreira, 1993). The issue, then, is whether a particular acoustic effect in speech is attributable to performance, or to the implementation of a linguistic (i.e., phonological) representation.

Psycholinguists tend to assume that prosodic effects arise at least in part because of factors related to performance during production. This can clearly be seen in the tendency to treat the notions of a 'performance unit' and an intonational phrase as essentially synonymous (Gee & Grosjean, 1983), and to use terminology which suggests that when a speaker for some reason needs to divide up an utterance for processing, the units that result will almost necessarily be prosodic constituents of some type (typically intonational phrases; e.g., see Watson & Gibson, 2004). On this view, then, some intonational phrases are planned top-down and generated from a discourse-semantic representation, but others are created on the fly, emerging when a speaker finds him- or herself needing to close off one processing chunk so that a new one can be initiated. Intonational phrases are then a byproduct of processing decisions.

To begin to address this question concerning the relationship between prosody and performance, I have chosen to adopt a strong position: I will assume that prosody and acoustic phenomena related to performance are distinct. Ultimately, I suspect this position will turn out to be too strong, for reasons that I will explore towards the end of the paper. But, as Ken Forster (1976) argued years ago, one is much more likely to uncover evidence for a theoretical distinction if one starts by assuming its existence. We therefore should give the strong hypothesis that the two are distinct phenomena a chance to prove itself before moving to the (arguably) weaker idea that they overlap to some significant extent.

The outline of this paper is as follows. First, I will describe the basic phenomena of interest. Next, I will review research that emerges from what I will term 'the algorithmic approach', where the goal is to try to generate a set of rules that predict the likelihood of a break at every word boundary within an utterance. The algorithmic approach was influential in the 1980s and has recently been resurrected by Watson and Gibson (2004). It has the essential property of treating prosody and performance phenomena as identical. Identifying the shortcomings of this approach should enable us to see in what ways the two need to be distinguished. In the third section I will consider evidence from experiments suggesting that the two are in fact separable and arise from different sources in the production system. These studies adopt an experimental approach: The goal is not to predict the likelihood of a break at every between-word location, but rather to manipulate the characteristics of linguistic material on the left side, right side, or both sides of a potential boundary, and then assess the effects on some specific set of dependent measures such as word and pause duration. This section also considers whether the algorithms predict the results found in those experiments. In the final section of the paper, I will describe the picture that emerges from the work that has been conducted up to this point, and I will lay out a plan of research that might help us to understand better not just prosody but also the general process of language production. I will also briefly consider the implications of these ideas for spoken language comprehension.

PHENOMENA AND DEFINITIONS

Let us begin with prosody (see Shattuck-Hufnagel & Turk, 1996, and Cutler, Dahan, and Van Donselaar, 1997, for more detailed reviews). Prosody can be divided into two main components: a metrical component and an intonational component (Bing, 1985; Ferreira, 2002; Inkelas & Zec, 1990; Samek-Lodovici, 2005; Selkirk, 1984; Warren, 1999; Zubizarreta, 1998). Metrical phonology is about sentence stress and duration – the sound features that cause an utterance to have a distinct rhythm (Goldsmith, 1990; Hayes, 1995; Liberman & Prince, 1977; Selkirk, 1984; Wagner, 2005). Consider this pair of sentences:

(1)
(a) Bill wants to go with Tom
(b) Tom wants to go with Bill

Assuming these sentences are said normally, so that information that is already established in the discourse is presented at the beginning of the sentence and new information is placed towards the end (the so-called given-new strategy; Haviland & Clark, 1974; see also Rochemont, 1986; Rooth, 1996; Selkirk, 1984), the word *Tom* would usually have a longer duration and would receive greater stress in (1a) than in (1b). The opposite pattern would hold for *Bill*. This tendency follows from the Nuclear Stress Rule (Chomsky & Halle, 1968), which states that the word at the end of a syntactic domain such as a clause is the most prominent.

The generalisation that syllables at the ends of major syntactic constituents tend to be louder and longer has been captured in a representational structure called the Metrical Grid (Prince, 1983; Selkirk, 1984)[1]. The metrical grid represents stress and, through the addition of silent positions in the grid, duration. The horizontal dimension of the grid indicates the organisation of syllables in time: the vertical dimension specifies degrees of stress. For example, consider (2).

[1] Metrical trees (Liberman & Prince, 1977) serve a similar purpose.

(2)

```
          5              X    5
X                        X         4
X    X         X         X         3
X    X         X         X         2
X    X    X    X    X    X         1
Bill wants to go with Tom
```

In Selkirk's (1984) theory of the metrical grid, the bottom level (level 1) provides a position for each syllable of the utterance, stressed or unstressed. Each position is referred to as a demibeat. The second level represents secondary word stress. Each position is referred to as a basic beat. The most prominent syllable of a content word receives a mark on the third level. Thus, this level represents main word stress. Levels 4 and above represent domain-end prominence rules, such as the rule that puts the strongest stress at the end of a phrase (the Nuclear Stress Rule), and these rules apply cyclically over increasingly larger syntactic domains. Because words at the ends of clauses are usually the most embedded (e.g., *Tom* in (2) occurs inside a prepositional phrase, a verb phrase, an infinitival phrase, and the entire sentence), they will receive marks at higher and higher levels, resulting in greater prominence. Note, however, that values in the metrical grid are interpreted relationally rather than absolutely; therefore, predictions are made only about relative prominence, not about precise phonetic values.

To account for domain-final lengthening and pausing, Selkirk proposed that silent positions are inserted into the grid by rules of silent demibeat addition (SDA). (Recall that demibeats are represented on level 1, the bottom level, of the grid.) The rules of SDA are the following:

Add a silent demibeat at the right edge of the metrical grid aligned with
(a) a word
(b) a word that is the head of a nonadjunct constituent
(c) a phrase
(d) a daughter phrase of S (daughter phrase of a clause)

Rules of SDA operate on the syntactic structure and contribute to the construction of a phonological representation – in this case, a metrical grid. For example, in (1), the word *Tom* would receive the greatest number of silent demibeats: it would get one for each of (a) through (d), and rule (c) would apply once for each phrase *Tom* belongs to (again, keeping in mind that the grid is not designed to predict absolute prominence, but instead specifies comparative stress and duration). Silent demibeats are associated with the syllable to the left and result in final lengthening. Selkirk hypothesised that a pause occurs if the syllable reaches the limits of its stretchability without absorbing all silent demibeats (but see Ferreira, 1993, for evidence against this hypothesis). Note that, according to this approach, the rules that determine how much lengthening and pausing will occur appeal to syntactic constituents such as phrases and clauses, not prosodic constituents such as phonological and intonational phrases (see Truckenbrodt, 1999, for a detailed discussion of this issue). This point will be relevant when we discuss the algorithmic approach in the next section of the paper. Notice too that the amount of lengthening and pausing that is predicted to occur is based entirely on the constituent structure that has already been created and not on any material that might be coming up in the utterance. Thus, the SDA approach assumes that lengthening and pausing are unrelated to the need to plan upcoming material. These effects arise entirely because of the linguistic properties of words to the left in the linguistic representation.

The second component of prosody is intonation, which refers to changes in pitch or fundamental frequency across an utterance. Tones of different types lead to distinct intonation contours (Beckman, Hirschberg, & Shattuck-Hufnagel, 2005; Bolinger, 1986; Ladd, 1996; Steedman, 2000a,b). For example, consider (3):

(3)
When the doctor scowled the patient became nervous
($_{\text{IPh}}$) ($_{\text{IPh}}$)

The precise relationship between syntactic structure and intonational phrasing is a major topic that space limitations do not allow us to consider in detail here (but see Steedman, 2000a,b, for a discussion). However, it is generally agreed that a clause boundary such as the one shown in (3) is an obligatory location for an intonational phrase boundary. Thus, the entire utterance would be spoken as two intonational phrases, with the boundary being marked with a sequence of tones commonly indicated with L-H% (meaning a low-tone followed by a sharp rise; Pierrehumbert, 1980; Pierrehumbert & Hirschberg, 1990). Intonational boundaries may also occur clause-internally, as in (*Mary knows*) (*the solution to the problem*). A major question for linguists and psycholinguists is what constitutes a permissible intonational unit, and what affects speakers' decisions about how to intonationally phrase an utterance (Ladd, 1996; Selkirk, 1984; Steedman, 2000a,b).

Now we turn to the less well-behaved domain of performance effects and disfluencies. Consider this example, spoken by a famous actor turned small-town mayor at a hearing of the United States Congress:

(4)
(Who in who in America) (gives these uh lawyers the) (to be the self-appointed vigilantes) (to uh to enforce the law)?

The speaker produced this utterance so there was a break (indicated with parentheses) after *America*, *lawyers the*, *vigilantes*, and *law.* In addition, unfilled pauses occurred before each *uh.* The utterance also contains standard disfluencies such as fillers (*uh*), repeats, repairs, false starts, and one outright abandonment (*gives these uh lawyers the – to be . . .*).

Disfluencies are clearly related to constituent structure. For example, when speakers produce repairs (e.g., *the blue dot I mean the red dot*), they typically backtrack to some type of phrasal boundary, so that the reparandum (the portion spoken in error) and the repair (the correction) together constitute a proper conjoined phrase (Levelt, 1983). In addition, pauses, hesitations, fillers, and repeats may occur anywhere in an utterance, but they are all more common at major syntactic boundaries than elsewhere (Goldman-Eisler, 1968). There is evidence that listeners use information from disfluencies to help them decide whether a constituent boundary should be postulated (Bailey & Ferreira, 2003; Ferreira & Bailey, 2004). However, there may be an important difference between pauses that arise due to disfluency and those that might occur because of the rules of SDA: The former seem related to the complexity of upcoming material, and the latter to the complexity of material already produced. In other words, by hypothesis, the likelihood and duration of a timing-based pause is related to the words and structure to the left of the boundary, and the likelihood and duration of a hesitation is related to material to the right of the boundary. At least, this is the story that emerges given the theoretical constructs presented thus far. The critical empirical question, of course, is whether the story is correct.

Before turning to the algorithms that have been proposed to account for the acoustic effects we have been discussing, we should note that if the approach that has been described here is correct, syntactic boundaries in real speech will be marked by a mixture of acoustic features created both by prosody and by planning. Therefore, at any particular sentential location, it will appear that both the length and complexity of previous material and the length and complexity of upcoming material influence measures such as pause time. But, critically, if there is a meaningful distinction between prosody and performance effects, with the former being attributable to left context and the latter to right context, then a pause at any single location can be ‘parsed’ into two subtypes, one attributable to implementation of the metrical grid, and the other to difficulties with planning new material. Alternatively, if prosody and performance phenomena overlap or are in fact the same thing, or if the processes that implement prosody and manage processing resources interact during language production, then left and right effects have some common sources in production.

A similar approach to understanding spoken language production was proposed by Dell, Burger, and Svec (1997), who studied speech errors generated when speakers attempted to say sequences such as *Gloria's Greek green gloves*. They distinguished between perseveration errors, which are due to a failure to deactivate material to the left, and anticipation errors, which arise due to planning of material to the right. Their work differs in some critical ways from the approach that is being developed here: for example, Dell et al. focus mainly on segmental aspects of production (e.g., pronouncing 'Greek' as 'Gleek'), and their dependent variable was error rate rather than amount of lengthening or pausing or the presence of intonational boundaries. Nonetheless, their model contains an important insight that is echoed in the present model of prosody and performance: that language production requires the cognitive and language systems to deal successfully with the past, the present, and the future. Specifically, Dell et al., argue that the past must be deactivated, the present must move into the system's current 'workspace', and the future must be prepared for. In addition, their experiments suggest that a successful system is one that emphasises the future over the past, as they observed that error rates decreased as anticipation errors began to dominate over perseverations. This idea that the past should be sacrificed for the future is one that has not yet been examined for the generation of prosody, but would be an intriguing project for future work.

Algorithmic approach

In the algorithmic approach, the goal is to develop a set of rules that can be used to predict the amount of some dependent variable that will occur at every potential between-word location. The idea is that the algorithm is like a grammar, in that it generates a representational structure for every sentence of the language, and that structure is then related to measures of interest. The measure may be pauses (Gee & Grosjean, 1983), phrase-final lengthening (Wagner, 2005), or the perceptual correlates of an intonational boundary (Watson & Gibson, 2004). The variable is assessed at every word boundary, and then simple correlations are computed between the predicted and obtained values (r represents the size of the correlation, and R^2 the proportion of variance in the dependent measure accounted for by the algorithm).

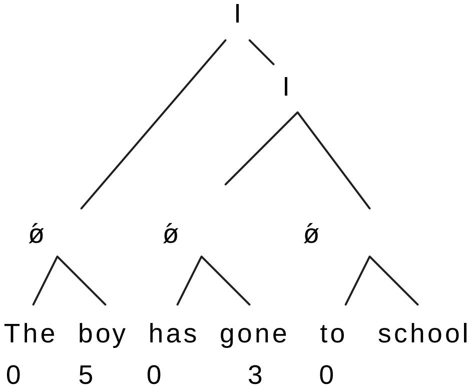

Figure 1. A performance structure generated by the Gee and Grosjean (1983) algorithm. 'I' = intonational phrase, and Ǿ = phonological phrase.

I will begin with the Gee and Grosjean (1983) algorithm (henceforth GG), which was designed to predict pauses over 200 ms in duration.[2] One goal was to account for what Gee and Grosjean consider the three main characteristics of 'pause structures', which may or may not coincide with the sentence's syntactic structure: The units separated by pauses are small (but see footnote 2), the overall pause structure is hierarchical, and the structure is more-or-less symmetrical. A major innovation of the algorithm was to suggest that it is not the syntactic level that is most appropriate for describing pauses, but rather what Gee and Grosjean refer to as the 'prosodic level'. (Notice that here their approach differs from Selkirk's (1984) rules for assigning prominence, which made reference to syntactic rather than prosodic constituency.) Gee and Grosjean argue for this level because they note that pauses often separate units that are not major syntactic constituents. For example, in the sentence *The kids left after I told them about the party*, the longest pause might occur between *left* and *after*. The major division in the sentence's syntactic structure, however, is between *The kids* and *left*. Thus, the units at the prosodic level do not necessarily correspond to the major units at the syntactic level. The two representational levels are not isomorphic.

The GG algorithm takes a sentence and constructs a 'pause structure' over it, as illustrated in Figure 1.The pause structure is similar to a syntactic

[2] Specifically, the GG algorithm was designed to predict the *duration* of pauses 200 ms or longer at every sentential location. Nonetheless, in Ferreira (1988) I argue that the algorithm can also be interpreted as predicting the probability of a pause, which in some circumstances makes the algorithm more plausible, since I found that people rarely paused more than once or twice in a single utterance, unless they were speaking at an unusually slow speech rate.

structure: Every position is dominated by a node, the nodes have labels, and the structure is a hierarchical tree. However, the pause structure differs from a syntactic structure in important ways. The positions in question are not lexical items, but rather the locations after them (where pauses may occur). The labels on the nodes dominating the pause locations consist of either a ǿ or I, where a unit dominated by ǿ corresponds to a phonological phrase, and I to intonational phrases. These units are not syntactic units, but units of the phonology, and have been widely discussed in the phonology literature (Jun, 2005; Nespor & Vogel, 1986; Selkirk, 1986; Zubizarreta, 1998). On the Gee and Grosjean definition, a phonological phrase consists of a series of function words up to and including a content word (e.g., *the boy*, *has gone*), while an intonational phrase is a larger unit over which fundamental frequency or pitch gradually drops. Phonological phrases are linked to form higher-level intonational phrases. The linking is accomplished through binary branching of the nodes into structures that take into account the sentence's syntactic structure and sometimes readjust it. The result is a representation similar to the metrical trees proposed by Liberman and Prince (1977).

The algorithm then generates pause values after each word using the pause structure tree. The intuition behind the pause assignment component of the algorithm is that short pauses, if any, occur within a phonological phrase, and long pauses occur after intonational phrases. In addition, the higher the intonational phrase is in a tree, the longer the pause will be. Thus in Figure 1, no pause is predicted to occur between *the* and *boy*, *has*, and *gone*, or *to* and *school*. A pause is predicted between *gone* and *to* and a larger one between *boy* and *has*. These predictions are generated by a complexity index that counts nodes and assigns pause locations a value based on the node count. The complexity index considers the nodes to the left and right of any pause location. According to GG, then, all the sentential material that both precedes and follows a word can affect the duration of the pause after it.

The algorithm was tested on data reported in Grosjean, Grosjean, and Lane (1979). In this earlier study, six participants read 14 sentences aloud. Each sentence was read six times: twice at a normal speech rate, twice at double the normal speech rate, and twice at half the normal speech rate. Pauses were defined as periods of no activity longer than 200 ms in a waveform for the utterance. Total pause time for a single sentence was calculated to obtain the proportion of pause time at each within-word location. Simple correlations were then computed between predicted and obtained pause duration. The mean correlation was .97; the lowest correlation for any sentence was .93, and the highest was .99. The algorithm thus appears to be a consistently accurate predictor of the pause structures of the sentences.

In addition, the apparent accuracy of the algorithm is attributable in part to the characteristics of the sentences on which it was tested. All 14 sentences were long and tended to have an obvious bisection point, and they were

TABLE 1
Sentences used to examine the Gee and Grosjean (1983) algorithm, with correlations between obtained pause durations and the predictions of both the GG algorithm and one based on rules of silent-demibeat addition

	Sentence	*GG correlation*	*SDA correlation*
1.	The priest seems terribly practiced and proficient at chess	.56	.91[b]
2.	The priest seems proficient at the chess game John bought for Bill	.93[c]	.64[a]
3.	The priest who lives next door seems proficient at the chess game John bought for Bill	.81[c]	.48
4.	Pete called up Anne on the telephone	.23	.76
5.	Pete called Anne up on the telephone	.24	.91[a]
6.	The girl that Mary said John talked to about the party for Bob collapsed on the floor	.63[b]	.77[c]
7.	The nurse that Mary contacted about Bill waited for the x-rays	.80[b]	.80[b]
8.	Who questioned John about Laurie after the dance?	.47	.64
9.	The lady wanted the dress on the rack for her sister	.86[c]	.94[c]
10.	The lady put the dress on the rack for her sister	.57	.64[a]
	Average Correlation	.61	.75

[a] $p < .05$; [b] $p < .01$; [c] $p < .001$.

fairly similar to each other in structure. I therefore tested the algorithm on ten new sentences that differed in a variety of ways (see Table 1): length (examples (1) –(5)), presence of a gap in the surface structure ((6) and (7)), and presence of arguments versus adjunct phrases ((8) – (10)). Productions were elicited in the same way as in Grosjean et al. (1979).

However, because it is critical to accurately separate pauses from stop closure durations at word edges, pauses were defined carefully based on an empirical criterion. The first step was to create a frequency distribution of all silent periods from 0 ms duration to infinity. This exercise revealed a bimodal distribution, dividing clearly at 80 ms. There was therefore no principled basis for using the GG value of 200 ms as the cutoff point for defining silence as a pause; instead, it appeared that pauses associated with producing stop consonants were those shorter than 80 ms, and the pauses related to the processes of interest (e.g., planning and prosody) were those longer than 80 ms.[3]

The results of this study are shown in Table 1. As can be seen in the second column, the GG algorithm did poorly on many sentences, and not

[3] The data were also analysed using the GG 200 ms the cutoff point, and results were similar.

TABLE 2
Correlations between the likelihood of placing an intonational boundary at every sentential location and (1) the probability of pausing; (2) the proportion pause time; (3) the Gee and Grosjean algorithm predictions; and (4) the silent demibeat predictions[4]

	Correlation			
Sentence	*(1)*	*(2)*	*(3)*	*(4)*
1.	.45	.42	.04	.18
2.	.54	.64[a]	.74[b]	.15
3.	.62[a]	.45	.41	.68[b]
4.	.41	.06	.80	.20
5.	–	–	–	–
6.	.59[a]	.76[c]	.53[a]	.82[c]
7.	.50	.68[a]	.73[a]	.82[b]
8.	.93[b]	.86[a]	.35	.70
9.	.72[a]	.64[a]	.55	.68[a]
10.	.94[c]	.92[c]	.59	.66[a]

[a] $p < .05$; [b] $p < .01$; [c] $p < .001$.

particularly well overall (average correlation was .61). An alternative algorithm that considered only the syntactic structure of material to the left of a potential pause location (essentially an implementation of Selkirk's Rules of SDA) performed somewhat better on average (average correlation was .75). It appears, then, that neither the GG algorithm nor a straightforward implementation of the rules of SDA predicts pauses accurately on a diverse (though admittedly small) set of sentences (see also Watson & Gibson, 2004, for similar findings and conclusions, although in their study the SDA-based algorithm did more poorly than GG).

Another aspect of the GG algorithm that should be evaluated is whether the intonational phrases that it generates are in fact perceived as such. To perform this analysis, three trained judges (the author and two phonologists) listened to each repetition of the ten sentences and indicated whether they heard an intonational break at any point. Locations on which all three judges agreed were classified as ones with an intonational boundary. The probability of placing an intonational boundary at every location in every sentence was then compared to the structures generated by the GG algorithm. The correlations are shown in Table 2. As can be seen, most correlations were modest, suggesting that the algorithm does not accurately predict the actual locations of major intonational phrase boundaries. This is a significant problem, because the purpose of the algorithm is to predict a

[4] No intonational breaks were perceived for any token.

sentence's intonational structure as well as its division into phonological phrases. Of course, the SDA-based algorithm also did not perform well, but recall that the rules of SDA are designed to predict the metrical properties of an utterance, not its intonational structure.

In summary, the GG algorithm captures some relevant facts about pausing, but it also suffers from some major weaknesses. One is that it assumes that both left and right context affect pauses. More worrisome is that this approach assumes a great deal of 'lookahead' (Levelt, 1989) on the part of the speaker; essentially, the entire sentence is known before the speaker begins planning its prosodic structure (indeed, the input to the algorithm is a sentence's entire lexical string). Most importantly, GG does not always predict pausing accurately, as indicated by the somewhat low correlations, and it does not accurately capture a sentence's perceived intonational phrasing either (see Table 2).

Now let us turn to the Watson and Gibson (henceforth, WG) algorithm (2004), which was designed to remedy the shortcomings in GG. The central concept is what they term 'Left/Right Boundary' (LRB) strength, which translates into the likelihood of an intonational phrase break at each between-word location. The model groups words into phonological phrases. These are defined in essentially the same way as in GG, as shown in (4), where parentheses separate phonological phrases:

(4)

```
(The judge) 4 (who the reporter) 2 (for the newspaper) 4
(ignored) 7 (fired) 1 (the secretary)
```

LRB strength is the sum of (a) the number of phonological phrases to the left of a particular location, and (b) the number of phonological phrases making up the constituent to the right of that same location, as long as that upcoming constituent is not an argument of the word to the left (thus, after the word *fired* in (4), the phrase *the secretary* is not counted because it is an argument of *fired*). In addition, a value of 1 is added if the word to the left terminates a phonological phrase. The numbers in example (4) are the predicted boundary strengths. The strongest boundary and the most likely location for an intonational phrase break is predicted to be after *ignored*, because four phonological phrases precede it (including itself), two follow it, and *ignored* marks the end of a phonological phrase (4 + 2 + 1). This location is the subject-verb phrase boundary, which also happens to be the syntactically most probable break point. In some circumstances, however, a different location will emerge, as was true for GG. Again, though, the WG algorithm will not allow a subject and verb to cluster together if the remainder would be an argument of that same verb (e.g., its direct object). A division after the verb could occur only if the following constituent were a modifier phrase of some type (e.g., a temporal phrase such as *last weekend*). A split before the modifier might occur if the subject of the sentence were short and modifier constituent long (defined in phonological phrase units), because both the GG and the WG algorithms try to create what are sometimes called ‘balanced sisters’ (Fodor, 2002): that is, prosodic constituents that are similar in size. But again, this tendency is attenuated in WG, because of the prohibition on intonational boundaries before postverbal arguments.

Also unlike the GG algorithm, the WG algorithm attempts to predict perceived intonational phrase boundaries (using the Tones and Breaks Indices (ToBI system; for a recent description, see Beckman et al., 2005), not pauses. In the study Watson and Gibson (2004) conducted to evaluate their algorithm, trained judges listened to sentences produced by naïve participants and indicated whether they heard an intonational break, and if so, where it was located. The predictions of their algorithm were then compared to GG, an alternative that considers only the left context (similar to one that implements the rules of SDA), and one that I proposed as part of my dissertation (Ferreira, 1988). Correlations between predicted and obtained breaks were similar for the GG, WG, and Ferreira (1988) algorithms (R^2 values were 76%, 74%, and 71% respectively) and much better than for the left-only algorithm (which accounted for only 39% of the variance). Because the first three R^2 values are not significantly different from each other, this study does not convincingly demonstrate the superiority of the WG algorithm. Another problem with the data Watson and Gibson present is that, as we saw earlier, the GG algorithm does a much better job at predicting pause durations than perceived intonational phrase boundaries (compare Tables 1 and 2). Thus, it is possible that if Watson and Gibson had measured pauses, the GG algorithm would have performed better. Recall from Section II that the metrical and intonational phonology are separate components of prosody, and it is probably true that pauses are more associated with timing than with intonation. Finally, it would be useful to know how well the WG algorithm performs on the same set of sentences used in the original Grosjean et al. (1979) study, again to make it easier to compare the accuracy of the two algorithms.

This last point brings up a major concern about the algorithmic approach in general, which is that the success of any algorithm depends in large part on the sentences selected to evaluate it. This is problematic, because thus far no principled basis for choosing sentences has been established, and the sets used across the different studies differ on many dimensions, making it difficult to determine what sentence properties are important. This, then, is one fundamental problem with the algorithmic approach: There is no agreed-upon method for choosing the stimuli on which to evaluate them. Another is that each algorithm thus far has considered only one dependent variable at a time (pause time or probability of a perceived intonational break), and the measures are not entirely comparable. An intonational phrase break is not always marked with a pause, and a pause may occur at locations other than intonational phrase boundaries (again, because pauses are probably more linked with timing than with intonation). In addition, it is not clear that separating pausing from phrase-final lengthening is defensible, because the two tend to be correlated (Cooper & Paccia-Cooper, 1980; Ferreira, 1993; Wagner, 2005) and indeed pauses are often not heard as such by naïve listeners but instead are perceived as syllable elongation (Martin, 1970).

Another useful way to test for the existence of phonological units would be to measure the operation of 'rules of external sandhi' (Kaisse, 1985; Nespor & Vogel, 1986; Selkirk, 1984). These are rules that may operate between words but only within a prosodic phrase. The classic example is French liaison (*le petit enfant*), but an example from English is the rule of across-word palatalisation, as in *Made you look*!. Most speakers of American English will palatalise the /d/ in *made*, but this process is blocked if a strong prosodic boundary intervenes, as in *Because of the way it was made your umbrella fell apart*. But except for a handful of studies (e.g., Cooper & Paccia-Cooper, 1980; Grabe & Warren, 1995), psycholinguists have not attempted to predict the locations where external sandhi rules are likely to occur or be blocked. It appears that this is an under-utilised tool for understanding prosody in language production. It would be interesting to know, for instance, whether the locations predicted by some algorithm to be the site of a long pause or an intonational phrase break are also places where sandhi rules fail to apply.

A further concern about the algorithmic approach as it has thus far been implemented is the use of correlations averaged over sentences to evaluate competing algorithms.[5] The problems are that, first, sentences differ in length, and this is unavoidable if one is to evaluate how comparatively well the algorithms perform for sentences that are short (or have some sort of short constituent) and those that are longer (or have longer constituents). Second, comparing correlations that have been averaged over sentences allows every sentence to have equal influence whether it varies a great deal on the dependent measure or just some small amount (in which case most of the variation is likely attributable to noise). Therefore, another approach might be to allow each between-word location to be a data point and NOT to calculate individual sentence-by-sentence correlations which then go into computing the overall correlation. However, this approach has a major shortcoming as well, which is that it makes it difficult to ascertain whether

[5] I am grateful to an anonymous reviewer for making this point.

sentences of some type are more problematic than others. Moreover, adopting this method would not allow data from current studies to be compared with those conducted before which made use of correlations averaged over sentences. What would be welcome, then, would be a more powerful and sophisticated technique for assessing model fits.

Finally, the reading task is not ideal for studying the issue we are concerned with, namely distinguishing genuine prosody from performance-related effects.[6] The main reason for having people read sentences rather than generate them spontaneously is that in most studies it is important to examine sentences with particular lexical, syntactic, and other properties, which have a low probability of being spontaneously produced (in the sense that any specific utterance is exceedingly rare). Another problem with spontaneous production is that if people have trouble generating the content, they may become highly disfluent, and hesitations due to production problems could easily 'swamp out' and even distort genuine prosodic effects. However, while it is true that speakers allowed to practice and prepare can read sentences fluently, it is also clear that they have the opportunity to plan more carefully and over much larger domains than in normal production. Indeed, if speakers produce speech incrementally (Ferreira & Swets, 2002; Griffin & Bock, 2000; Levelt, 1989; Smith & Wheeldon, 2001), then it is not clear that they have planned much beyond the current word, and so algorithms that assume effects of material far to the right of a potential boundary location are psychologically unrealistic. Finally, the reading task could very well distort even the prosody that is produced, because people do not read the way they naturally talk. Consider that when you listen to a radio broadcast, you can often tell when a speaker has switched from talking to reading from a prepared text. When people are asked to read a sentence 'naturally', they probably try to generate idealised prosodic forms, which may be different from the prosody they would produce naturally (Albritton, McKoon, & Ratcliff, 1996). Thus, what is needed is a task that allows the experimenter to control exactly what the speaker says in fairly complex utterances, does not require so much planning that highly disfluent renditions are likely to occur, and that also preserves at least some features of regular speech. Unfortunately, no such task currently exists.

EXPERIMENTS ON PROSODY AND PERFORMANCE

In this section I will summarise the results of some experiments I published years ago (Ferreira, 1988, 1991, 1993), in which the fundamental goal was to

[6] The shortcomings of reading tasks that will be identified are true also of the version of the reading task created by Watson and Gibson (2004).

distinguish prosody from planning effects, and to assess effects of left and right context independently. To make clear how to compare these studies with the ones conducted to assess the algorithms, it is important to highlight some features of the approach that I adopted.

The first is that I assessed prosody by measuring word AND pause durations. I did not measure perceived intonational boundaries because it was not clear at that time what information sources affect people's impression that they hear or do not hear a boundary (the work was done prior to the publication of ToBI), and I wanted a measure that was more implicit and could pick up on features not available to conscious introspection. I also decided to measure both word and pause durations because the hypothesis I was testing was that pauses were of at least two types (see Butterworth, 1980, for similar ideas): timing-based pauses inserted because of the metrical grid, and planning-based pauses inserted when the speaker needs extra time to organise upcoming material. The former, I hypothesised, were prosodic, and the latter, planning related. And because timing can be implemented using both phrase-final lengthening and pausing, I decided it was important to measure both at any particular boundary. Timing-based pauses and lengthening should co-occur according to the metrical grid model described in Selkirk (1984), and also based on the pioneering work of Cooper and his colleagues (Cooper & Paccia-Cooper, 1980). Planning-based pauses and sentence initiation times (which were also recorded) should co-occur but should not be related to lengthening, because initiation times and pauses both reflect the time it takes for speakers to plan a stretch of speech (not perfectly, of course, since speakers can also plan while they articulate; Ferreira & Swets, 2002; Levelt, 1989).

Second, a fundamental goal of the work was to assess whether left or right context affected word and pause durations, and to determine whether any such effects were a function of prosodic or syntactic constituency. Thus, the sentences that people were asked to produce were carefully varied to have either identical prosodic structures but different syntactic constituencies, or vice versa. It was therefore critical to control precisely what people said. For syntactic constituency to vary but not any relevant phonological features of the sentences, factors such as number of syllables and stress patterns had to be identical across conditions. For example, in the first experiment that I will describe, participants said sentences whose subjects included prenominal modifiers, a postnominal prepositional phrase, or a relative clause (e.g., *the enthusiastic band/the pianist in the band/the man who was in the band*). The object varied in exactly the same way (but with different content, of course). Importantly, though, the three versions had the same number of syllables and the same pattern of strong and weak syllables. They differed in number of words, but prior work conducted by Sternberg and colleagues (Sternberg, Monsell, Knoll, & Wright, 1978) had already demonstrated that number of words as they are normally defined does not affect variables such as initiation time and pause duration; what mattered in their work was length in prosodic words, and on that measure the conditions were equated. The word at the end of the relevant phrase was always one syllable in length and relatively easy to segment in a waveform (e.g., *band*, *crowd*).

Of course, to get people to say sentences that were this tightly controlled, the content of the sentences had to be provided to speakers. At the same time, I did not want to use the reading task for the reasons given earlier. I decided to use instead a variant of the memorisation task developed by Sternberg et al. (1978). The paradigm was as follows. On any trial, participants (who were seated in front of a computer with an attached voice-activated relay apparatus, a monitor, and a tape-recorder) saw a prompt asking them to push a button when they were ready to begin. Then a sentence such as *The enthusiastic band pleased the senator in the crowd* appeared and stayed on the screen until the participant hit a button to indicate that he or she had memorised it. At that point, the prompt *What happened?* appeared, and the participant's task was to answer the question with the memorised sentence. A variable delay was employed, following Sternberg et al., to ensure that participants could not anticipate when it would be time to begin to speak. Initiation times were recorded via the voice-key device, and all responses were recorded for later digital analysis. The dependent measures were duration of the subject-final word (*band* in the example), the pause after the subject-final word (which was also immediately before the sentence's main verb), the pause after the main verb, and the sentence-final word (*crowd* in the example), as well as sentence initiation times.[7]

In Ferreira (1988, 1991), the syntactic complexity of the subject and object were independently varied, as shown in (4).

(4)

The enthusiastic band		the very impatient crowd
The pianist in the band	pleased	the senator in the crowd
The man who was in the band		the girl who was in the crowd

The results (described in detail in Ferreira, 1991) were straightforward. First, initiation times were longer the more complex the subject of the sentence, and object complexity had no effect. Second, pause times before the verb (i.e., after the final word of the subject) were longer the more complex the

[7] Memorisation times were also recorded, but did not vary systematically with the independent variables (see Ferreira, 1991).

object, and subject complexity had no effect. Pauses *after* the verb were rare, and their probability and duration were unrelated to the independent variables (more recently, Elordieta, Frota, & Vigario, 2005 have found evidence that in both Spanish and Portuguese sentence-internal breaks almost always occur between the subject and main verb). Finally, word durations were entirely unaffected by the manipulations (Ferreira, 1988) – the duration of *band* did not vary with the syntactic complexity of the subject, and the duration of *crowd* did not vary with the syntactic complexity of the object.

This set of findings as a whole suggests that the complexity of upcoming material affects pause durations, but these were not timing-based pauses – that is, these were not pauses created to implement a metrical representation. This conclusion follows from the fact that the pauses patterned with sentence initiation times, and from the lack of any phrase-final lengthening. It appears that most participants produced these utterances by dividing them into two syntactically defined units,[8] one consisting of the subject, and the other the verb phrase. The more complex the structure, the more planning time was needed, resulting in increased initiation times and pause durations. These findings support a fairly incremental model of language production, as initiation times were unaffected by object complexity. It appears that the system did not plan beyond more than a few words and not much past the current phrase (Ferreira & Swets, 2002; Griffin & Bock, 2000). The absence of word lengthening suggests that the metrical grids across all three conditions did not differ, which implies that prosodic constituency is the relevant level of the grammar for assigning timing. The results of another experiment support this account (Ferreira, 1993). Participants produced sentences such as either *The friend of the cop infuriated the boyfriend of the girls* or *The cop who's a friend infuriated the boyfriend of the girls.* Here it was found that the duration of *cop* was longer in the first case than in the second, consistent with the idea that word durations are affected by prosodic constituency (in the first sentence, the word *cop* occurs at the end of a prosodic constituent; in the second, it is located in the middle of the same constituent). In addition, pause durations did pattern with word durations, indicating that these were timing-based pauses, not the planning-based pauses elicited in the other experiment.

It appears, then, that left and right context have markedly different effects on word and pause durations. Upcoming material affects pause but not word durations. Initiation times are also affected, because pauses are essentially initiation times for sentence-internal constituents. Left context is not related to performance. In contrast, left context affects prosodic processes such as phrase-

[8] The units could also be described as semantic, because the subject-predicate boundary is also a major semantic division.

final lengthening and the pauses that co-occur with lengthening.[9] Right context is irrelevant. Thus, returning to the algorithmic approach, it is clear that algorithms should distinguish planning from timing. They appear to work – that is to predict some dependent variable accurately – to the extent that they capture both timing and planning (and as we saw from the reported correlations between predicted and obtained values, they do not perform especially well). But even when they predict successfully, they still present a misleading picture of how the production system works, because they do not distinguish between planning and timing, and they contribute to the misunderstanding that the prosodic and performance effects always have the same source. As careful experimental studies show, the two sources are often different.

Let us now consider whether either the GG or the WG algorithm can account for the results of these experiments. The GG algorithm does poorly, as pointed out in Ferreira (1991), because it predicts that the location of the main pause in a sentence will shift from before the verb to after it when the subject is simple and the object complex (in an attempt to create two equal-sized prosodic units). But as the findings for both pause duration and pause probability show (Ferreira, 1991), speakers almost always pause before the verb; pauses after the verb are rare and are essentially randomly distributed. The problem with the GG algorithm is that it does not take the strength of the subject-predicate boundary seriously enough, and thus allows the main break in a sentence to shift fairly easily (depending on complexity) from before the matrix verb to after it. As Watson and Gibson (2004) point out, a pause immediately before a verb's object is not very likely.

The WG algorithm is more successful than GG, in part because it is less likely to predict intonational breaks after the matrix verb. This is because WG forbid placing an intonational phrase boundary before a verb's internal arguments. However, although this prohibition seems correct for direct objects, and perhaps also for cases in which a verb has two internal arguments (e.g., (*John put*) (*the book on the table*) does seem ill-formed), it also appears to be too strong in some cases. For example, consider verbs that take a clause as an argument, as in *Mary believes that the appeal will be successful.* In this case, a break after the verb *believes* seems perfectly acceptable. Notice that the experiments summarised in this section did not examine cases of clausal arguments, and so this idea needs to be empirically evaluated.

Another reason the WG algorithm does better than GG is that it is more incremental: Only the upcoming constituent to the right of a boundary is considered (and only if it is a non-argument, as already noted), not the entire remainder of the sentence. Still, the algorithm does not perform perfectly (as

[9] Ferreira (1993) describes in detail in what way lengthening and pausing co-occur, and provides a mathematical model describing the trade-off between the two processes for implementing a metrical grid.

the R^2 values indicate). In addition, consider what is predicted to happen when the subject is simple and the object complex, as in *The enthusiastic band pleased the girl who was in the crowd.* This sentence would be assigned LRB values as follows:

(5)

```
The enthusiastic band) pleased) the girl) who was in the crowd
                     1+ 3       0+ 0    1+ 1
                      + 1        + 1     + 1
```

The most likely place for a break is still predicted to be after the subject, which is consistent with the experimental results. However, notice that the second most likely location is predicted to be after *girl*, right in the middle of the sentential object. But this phrasing would be extremely ill-formed, as it would combine the subject and matrix verb with a piece of the object, and would leave the relative clause in the object isolated as a separate intonational constituent. This structure violates all versions of the Sense Unit Condition (Selkirk, 1984; Steedman, 2000b), which in essence states that intonational phrases should be semantic units (see the cited papers for more formal definitions). Thus, there is reason to question whether this location is indeed the second most likely for an intonational break. Moreover, if the verb in the relative clause were a content word (e.g., *feared*) rather than *was in*, the location after *girl* would have an even higher value, because there would be two phonological phrases in the upcoming constituent. The value before the verb would still be 5 but the value after *girl* would increase to 4, making the two locations similar in probability of a break, and both would be much more likely to be the site of a break point than any other locations within the sentence. (Still, it is important to emphasise that the WG algorithm correctly predicts whether a break is more likely before or after the sentence's main verb, and this is an important fact for any algorithm to get right.)

What is most problematic, though, is that neither the GG nor the WG algorithm captures the central point that emerges from the experiments summarized here (Ferreira, 1991), and that is that pauses caused by material on the left and material on the right may have fundamentally different sources in the production system. Recall that pauses increased with upcoming object complexity just as initiation times increased with upcoming subject complexity, but word durations did not co-vary. Pauses were affected only by material to the right, not by the characteristics of the material to the left. This pattern suggests that the pauses were in fact hesitations, inserted for the purposes of planning upcoming material. Moreover, the more syntactically complex the upcoming phrase, the longer the pause before it. In contrast, the WG algorithm predicts no difference between the prepositional phrase and relative clause modifier conditions (e.g., *the senator in the crowd/the girl who was in the crowd*), because the number of phonological phrases is the same in the two conditions. A critical finding from the Ferreira (1991, 1993) experiments, though, is that upcoming syntactic (or semantic) complexity is what affects planning, not complexity defined in terms of phonological units.

WHAT ALL THIS MEANS

Nothing that was presented in the preceding section describing experiments on prosody and performance implies that it is wrong to develop and evaluate algorithms for predicting language production behavior. The criticisms all have to do with the specific versions that have been proposed and the ways they have been evaluated. Thus, the risk of the algorithmic approach is that the reasons for any algorithm's success or failure will not be carefully assessed. The GG algorithm appeared to be essentially perfect, but once it was tested on a new set of sentences, it became apparent that it is seriously flawed. What is necessary is to understand clearly what principles the algorithm is instantiating, and make sure to choose stimuli that allow those principles to be straightforwardly tested.

For example, the model that I am proposing assumes that prosody is distinct from performance effects. An algorithm could be developed that assumes precisely these principles. It could work something like this (see Ferreira, 1993, for more details): The production system would build some amount of the upcoming semantic-syntactic structure before phonological encoding (Ferreira, 2000). Assuming moderate incrementality, no more than one or two words or a single phrase would be anticipated (as in WG). The length and complexity of that semantic-syntactic structure would be positively correlated with initiation times as well as pause durations. At the same time, as the utterance was being prepared for production, a prosodic structure would also be created, and it would be used to determine the amount of phrase-final lengthening and pausing for a particular syllable. Thus, the algorithm would look at prosodic constituents to the left of a boundary but it would consider semantic-syntactic constituents to the right of that same boundary, and it would clearly distinguish between pauses attributable to a linguistic representation like a metrical grid from those caused by the need to plan. In addition, to test the accuracy of the algorithm, a tightly controlled set of sentences should be used, along the lines of the stimuli employed in the experimental studies summarised earlier. But there is no reason why the algorithm could not be used to generate predictions. Thus, a hybrid of the experimental and algorithmic approaches is probably the optimal research strategy. The algorithms generate precise predictions, and the experiments allow the predictions to be tested logically and systematically, with a principled set of stimuli.

Another approach to evaluating prosody and performance in language production would be to design experiments to make planning at a particular location difficult, in order to assess the prosodic structure that is observed. A shortcoming of most previous work, including my own, is that the difficulty of planning has been manipulated by varying structural characteristics such as number of words or syntactic complexity. Unfortunately, depending on one's theory of the mapping from syntax to prosody, these manipulations are potentially representationally ambiguous. A longer phrase must contain more syntactic nodes, but it will often contain more prosodic words as well (Watson & Gibson, 2004, make this same point). And longer phrases tend to be semantically more complex. Thus, the various representational systems are difficult to disentangle. But imagine that a lexical variable such as word frequency was manipulated. Specifically, consider a sentence in which the main verb was frequent, and another in which the main verb was rare. Imagine further that the verbs were matched on number of syllables, stress pattern, and so on. If a break turned out to be more common before the verb when it was more difficult to retrieve (i.e., less frequent), that would suggest an influence related to planning, not prosody, as no theory of prosody takes into account lexical frequency. And if the break were carefully examined to assess whether it had the features of a genuine intonational phrase boundary (e.g., the right pattern of tones), it would be possible to determine whether planning induces prosodic boundaries (see Gahl & Garnsey, 2004, confirming that the probability of a given syntactic constituent given a lexical item affects pronunciations and word durations).

If it is true that prosody and performance have different sources in production, there are interesting implications for language comprehension. A great deal of energy has been devoted to trying to discover whether prosodic information can be used by listeners to help them build syntactic and semantic structures (Clifton, Carlson, & Frazier, 2002, 2006; Eckstein & Friederici, 2005; Kjelgaard & Speer, 1999; Price, Ostendorf, Shattuck-Hufnagel, & Fong, 1991; Schafer, Carlson, Clifton, & Frazier, 2000; Weber, Grice, & Crocker, 2006). This approach, though, assumes a fairly tight link between prosody and syntax. But if people speaking naturally tend to be disfluent, and if essentially any content word may be preceded by a pause due to word retrieval problems, then listeners might be frequently misled into postulating absurd syntactic structures. Recall that although disfluencies such as hesitation pauses tend to be most common at the left edge of syntactic constituents, the second most likely location is one word in, especially after an initial determiner (Maclay & Osgood, 1959). But should a listener postulate a right syntactic bracket after *the*, given that the result could not be grammatical? Of course, the answer is no, but why not? Some determiners can constitute their own syntactic phrases

– *that*, for instance – so it is not true that the grammar forbids this possibility. Perhaps, then, the system discounts the acoustic effects around the determiner because they clearly signal disfluency rather than prosody. But then this idea is precisely what this paper is proposing: that to some extent the two systems are distinct, and moreover, that the comprehension system can even tell them apart. Recently, evidence has been uncovered suggesting that boundaries that might be attributed to planning are not as likely to be treated by the comprehension system as cues to syntactic structure compared to those that are more obviously related to prosody (Clifton, Carlson, & Frazier, 2006). Thus, if the parser can distinguish between prosody and performance effects, perhaps it is because the speaker produces different acoustic effects for prosody and for performance.

Again, though, it is important to appreciate how interesting the hypothesis is that prosody and performance are in fact related. This idea suggests an approach that might be termed *naturalised phonology*, where prosody is ultimately grounded in the psychological system that supports communication (as argued by Lieberman, 1984). On this view, perhaps all boundaries are ultimately attributable to performance, but some have become grammaticised because they are so frequent and therefore listeners come to expect them. For example, perhaps the boundary between two clauses is an obligatory location for an intonational phrase break because almost all speakers need time to organise an upcoming clause. The boundary might tend to be more pronounced when the clause is long because more time and effort is required for longer and more complex sequences, but even if the upcoming clause is short a boundary will be included. This regularity could be a diachronic consequence of speakers' regular tendencies – in other words, the grammar comes to include a pattern that is almost universal given the limitations of the performance system.

In addition, the proposal I have been making is that effects of material on the left are due to prosody, and effects of material on the right are due to performance – planning, specifically. But consistent with the enterprise of naturalising phonology and prosody in particular, it is possible that left-based effects are also due to the system's need to recover from having generated difficult or complex phrases (either phonological, syntactic, or semantic) (Cooper & Paccia-Cooper, 1980; Watson & Gibson, 2004). In other words, in other domains of action, pauses occur when the system needs to plan, but they also arise because of the need to rest after having executed a demanding sequence. On this view, phrase-final lengthening and pauses correlated with lengthening are at least in part attributable to recovery from material that has already been produced, and initiation times and planning pauses are due to anticipation of future material. An intriguing possibility is that, as Dell et al. (1997) observed, successful production involves emphasising the future over the past (the right over the left), suggesting that when people are fluent, they might pause less due to recovery than to planning. This is an important idea to investigate in future work.

Finally, models of production need to take more seriously the question of how speakers choose a prosodic structure for an utterance. A large literature now exists addressing how speakers choose syntactic forms (see Christianson & Ferreira, 2005, for references and discussion). This work suggests that syntactic choices emerge from two distinct processes. First, a speaker might produce a passive because he or she has a semantic intention that translates into a structure in which a theme or patient needs to be topicalised. The passive, then, is planned early in the production process, when the message level is formulated. Alternatively, a speaker might find that for a variety of reasons the patient or theme is highly activated or available (e.g., it has been primed), and to take advantage of this situation and promote fluency and incremental production, the system might plug that activated concept into the subject position of a syntactic tree, resulting in a passive (Bock, 1986). On this latter view, the passive is a byproduct of psychological processing; it was not planned, it emerged. Imagine that a similar story holds for prosody. Some prosodic structures might be planned early, as part of the speaker's organisation of information. In other words, to convey the semantic information associated with various intonational tunes (Steedman, 2000a), the speaker might formulate a particular intonational structure during message level planning. This would be comparable to planning a passive based on the organisation of given and new information. But speakers might also find themselves in the middle of producing a long or difficult constituent and discover that they need time to plan upcoming material. In this circumstance, perhaps the production system takes a break by inserting a prosodic boundary (as argued by Watson and Gibson, 2004, for instance). Here, the prosodic form of the entire utterance would emerge as a by-product of the system's attempts to manage its finite processing resources.

In conclusion, as stated at the outset, the aim of the paper was to distinguish prosody from acoustic effects attributable to performance. But this is really the superficial goal of the work; the true goal is to try to sharpen the debate by highlighting a distinction that is likely important to consider at every stage as we continue working on language production. As I have stated several times, I do not think the strong view that prosody and performance are entirely different will ultimately turn out to be correct, but that is not important. What is critical, I believe, is to foster more research on prosody and disfluency in language production, and to put at least as much effort into understanding the sounds of utterances as their syntax. If this happens, we will greatly expand our understanding of all language systems.

REFERENCES

Albritton, D., McKoon, G., & Ratcliff, R. (1996). Reliability of prosodic cues for resolving syntactic ambiguity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 714–735.

Altmann, E. M. (2004). Advance preparation in task switching: What work is being done? *Psychological Science*, *15*, 616–622.

Bailey, K. G. B, & Ferreira, F. (2003). Disfluencies influence syntactic parsing. *Journal of Memory and Language*, *49*, 183–200.

Beckman, M. E., Hirschberg, J., & Shattuck-Hufnagel, S. (2005). The original ToBI system and the evolution of the ToBI framework. In S. Jun (Ed.), *Prosodic typology: The phonology of intonation and phrasing* (pp. 9–54). New York: Oxford University Press.

Bing, J. M. (1985). *Aspects of English prosody*. New York, NY: Garland Publishers.

Bock, J. K. (1986). Meaning, sound, and syntax: Lexical priming in sentence production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *12*, 575–586.

Bolinger, D. (1986). *Intonation and its parts: melody in spoken English*. Stanford, CA: Stanford University Press.

Butterworth, B. (1980). *Language production*. New York: Academic Press.

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

Christianson, K., & Ferreira, F. (2005). Conceptual accessibility and sentence production in a free word order language (Odawa). *Cognition*, *98*, 105–135.

Clifton, Jr., C., Carlson, K. & Frazier, L. (2002). Informative prosodic boundaries. *Language and Speech*, *45*, 87–114.

Clifton, Jr., C., Carlson, K., & Frazier, L. (2006). Tracking the what and why of speakers' choices: Prosodic boundaries and the length of constituents. *Psychonomic Bulletin and Review*, *13*, 854–861.

Cooper, W. E., & Paccia-Cooper, J. (1980). *Syntax and speech*. Cambridge, MA: Harvard University Press.

Cutler, A., Dahan, D., & Van Donselaar, W. A. (1997). Prosody in the comprehension of spoken language: A literature review. *Language and Speech*, *40*, 141–202.

Dell, G. S., Burger, L. K., & Svec, W. R. (1997). Language production and serial order: A functional analysis and a model. *Psychological Review*, *104*, 123–147.

Eckstein, K., & Friederici, A. (2005). Late interaction of syntactic and prosodic processes in sentence comprehension as revealed by ERPs. *Cognitive Brain Research*, *25*, 130–143.

Elordieta, G., Frota, S., & Vigario, M. (2005). Subjects, objects, and intonational phrasing in Spanish and Portuguese. *Studia Linguistica*, *59*, 110–143.

Ferreira, F. (1988). *Planning and timing in sentence production: The syntax-to-phonology conversion*. Unpublished dissertation, University of Massachusetts, Amherst, MA.

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. (2000). Syntax in language production: An approach using tree-adjoining grammars. In L. Wheeldon (Ed.), *Aspects of language production* (pp. 291–330). Cambridge, MA: MIT Press.

Ferreira, F. (2002). Prosody. In *Encyclopedia of cognitive science*. London: Macmillan Reference Ltd.

Ferreira, F., & Bailey, K. G. D. (2004). Disfluencies and human language comprehension. *Trends in Cognitive Science*, *8*, 231–237.

Ferreira, F., & Swets, B. (2002). How incremental is language production? Evidence from the production of utterances requiring the computation of arithmetic sums. *Journal of Memory and Language*, *46*, 57–84.

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

Forster, K. I. (1976). Accessing the mental lexicon. In R. J. Wales & E. Walker (Eds.), *New approaches to language mechanisms* (pp. 257–287). Amsterdam: North-Holland.

Gahl, S., & Garnsey, S. M. (2004). Knowledge of grammar, knowledge of usage: Syntactic probabilities affect pronunciation variation. *Language*, *80*, 748–775.

Gee, J. P., & Grosjean, F. (1983). Performance structures: A psycholinguistic and linguistic appraisal. *Cognitive Psychology*, *15*, 411–458.

Goldman-Eisler, F. (1968). *Psycholinguistics: Experiments in spontaneous speech*. New York: Academic Press.

Goldsmith, J. A. (1990). *Autosegmental and metrical phonology.* New York: Blackwell Publishers.

Grabe, E., & Warren, P. (1995). Stress shift: do speakers do it or do listeners hear it? In B. Connell & A. Arvaniti (Eds.), *Phonology and phonetic evidence: Papers in laboratory phonology IV* (pp. 95–110). Cambridge, UK: Cambridge University Press.

Griffin, Z., & Bock, K. (2000). What the eyes say about speaking. *Psychological Science*, *11*(4), 274–279.

Grosjean, F., Grosjean, L., & Lane, H. (1979). The patterns of silence: Performance structures in sentence production. *Cognitive Psychology*, *11*, 58–81.

Haviland, S. E., & Clark, H. H. (1974). What's new? Acquiring new information as a process in comprehension. *Journal of Verbal Learning and Verbal Behavior*, *13*, 512–521.

Hayes, B. (1995). *Metrical stress theory.* Chicago, IL: University of Chicago Press.

Inkelas, S., & Zec, D. (1990). *The phonology-syntax connection*. Chicago, IL: University of Chicago Press.

Jun, S. (2005). *Prosodic typology: The phonology of intonation and phrasing*. New York: Oxford University Press.

Kaisse, E. M. (1985). *Connected speech: The interaction of syntax and phonology.* New York: Academic Press.

Kjelgaard, M. M., & Speer, S. R. (1999). Prosodic facilitation and interference in the resolution of temporary syntactic closure ambiguity. *Journal of Memory and Language*, *40*, 153–194.

Ladd, R. D. (1996). *Intonational phonology.* New York: Cambridge University Press.

Levelt, W. J. M. (1983). Monitoring and self-repair in speech. *Cognition*, *14*, 41–104.

Levelt, W. J. M. (1989). *Speaking*. Cambridge, MA: MIT Press.

Liberman, A., & Prince, A. (1977). On stress and linguistic rhythm. *Linguistic Inquiry*, *8*, 249–336.

Lieberman, P. (1984). *The biology and evolution of language.* Cambridge, MA: Harvard University Press.

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

Martin, J. G. (1970). On judging pauses in spontaneous speech. *Journal of Verbal Learning and Verbal Behavior*, *9*, 75–78.

Nespor, M., & Vogel, I. (1986). *Prosodic phonology.* Dordrecht, the Netherlands: Foris Publications.

Pierrehumbert, J. B. (1980). *The phonology and phonetics of English intonation.* Ph.D. dissertation, Massachusetts Institute of Technology, Boston, MA.

Pierrehumbert, J. B., & Hirschberg, J. (1990). The meaning of intonation contours in the interpretation of discourse. In P. R. Cohen, J. Morgan, & M. E. Pollack (Eds.), *Intentions in communication* (pp. 271–311). Cambridge, MA: MIT Press.

Price, P. J., Ostendorf, M., Shattuck-Huffnagel, S., & Fong, C. (1991). The use of prosody in syntactic disambiguation. *Journal of the Acoustical Society of America*, *90*, 2956–2970.

Prince, A. (1983). Relating to the grid. *Linguistic Inquiry*, *14*, 19–100.

Rochemont, M. S. (1986). *Focus in generative grammar.* Amsterdam: J. Benjamins.

Rooth, M. (1996). Focus. In S. Lappin (Ed.), *The handbook of contemporary semantic theory* (pp. 271–297). Oxford: Blackwell.

Samek-Lodovici, V. (2005). Prosody-syntax interaction in the expression of focus. *Natural Language and Linguistic Theory*, *23*, 687–755.

Schafer, A., Carlson, K., Clifton Jr., C., & Frazier, L. (2000). Focus and the interpretation of pitch accent: Disambiguating embedded questions. *Language and Speech*, *43*, 75–105.

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

Selkirk, E. O. (1986). On derived domains in sentence phonology. *Phonology Yearbook*, *3*, 371–405.

Shattuck-Hufnagel, S., & Turk, A. (1996). A prosodic tutorial for investigators of auditory sentence processing. *Journal of Psycholinguistic Research*, *25*, 193–247.

Smith, M., & Wheeldon, L. (2001). Syntactic priming in spoken sentence production – an online study. *Cognition*, *78*, 123–164.

Steedman, M. (2000a). Information structure and the syntax-phonology interface. *Linguistic Inquiry*, *31*, 649–689.

Steedman, M. (2000b). *The syntactic process*. Cambridge, MA: MIT Press.

Sternberg, S., Monsell, S., Knoll, R. L., & Wright, C. E. (1978). The latency and duration of rapid movement sequences: Comparisons of speech and typewriting. In G. E. Stelmach (Ed.), *Information processing in motor control and learning* (pp. 117–152). New York: Academic Press.

Truckenbrodt, H. (1999). On the relation between syntactic phrases and phonological phrases. *Linguistic Inquiry*, *30*, 219–255.

Wagner, M. (2005). *Long distance effects on prosody.* Poster presented at the 18th Annual Meeting of the CUNY conference on Human Sentence Processing. Tucson, AZ, USA.

Warren, P. (1999). Prosody and language processing. In S. Garrod & M. Pickering (Eds.), *Language processing* (pp. 155–188). Hove, UK: Psychology Press Ltd.

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

Weber, A., Grice, M., & Crocker, M. (2006). The role of prosody in the interpretation of structural ambiguities: A study of anticipatory eye movements. *Cognition*, *99*, B63–B72.

Zubizarreta, M. L. (1998). *Prosody, focus, and word order*. Cambridge, MA: MIT Press.