

Exploring the Similarity Between Implicit and Explicit Prosody: Prosodic Phrasing and Individual Differences

Language and Speech

1–27

© The Author(s) 2020

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/0023830920972732

journals.sagepub.com/home/las



Jason Bishop 

City University of New York (College of Staten Island & The Graduate Center), USA

Abstract

In recent years, work carried out in the context of the implicit prosody hypothesis (IPH) has called into question the assumption that implicit (i.e., silently generated) prosody and explicit (overtly produced) prosody are similar in form. Focusing on prosodic phrasing, the present study explored this issue using an individual differences approach, and using methods that do not rely on the sentence comprehension tests characteristic of work within the IPH program. A large group of native English speakers participated in a production experiment intended to identify individual differences in average prosodic phrase length, phonologically defined. We then explored whether these (explicit) prosodic differences were related to two other kinds of variation, each with a connection to implicit prosody. First, we tested whether individual differences in explicit prosodic phrase length were predicted by individual differences in working memory capacity, a relationship that has been established for implicit prosody. Second, we explored whether participants' explicit prosodic phrase lengths were predictive of their behavior in a silent-reading task in which they had to identify their own implicit prosodic groupings. In both cases, the findings are argued to be consistent with a similarity between explicit and implicit prosody. First, participants with higher working memory capacity (as estimated by reading spans) were associated with longer prosodic phrases. Second, participants who produced longer explicit prosodic phrases in speech tended to report generating longer prosodic phrases in silent reading. Implications for the nature of implicit prosody, and how it can be studied, are discussed.

Keywords

implicit prosody, individual differences, working memory, speech production

Corresponding author:

Jason Bishop, City University of New York, 365 Fifth Avenue, New York, NY 10016, USA.

Email: jbishop@gc.cuny.edu

Introduction

1.1 Background

Nearly 20 years ago, Fodor (1998, 2002) proposed an understanding of prosody's role in sentence processing that she called the IPH, formulated as in (1) and crucially based on the working assumptions in (2):

- (1) The IPH (Fodor, 2002, p. 113)

In silent reading, a default prosodic contour is projected onto the stimulus, and it may influence syntactic ambiguity resolution. Other things being equal, the parser favors the syntactic analysis associated with the most natural (default) prosodic contour for that sentence.

- (2) Working Assumptions of the IPH:
- a. The implicit prosody projected onto the sentence in reading will be identical to the explicit (i.e., overtly spoken) prosody for that sentence in a comparable context.
 - b. The comprehender will be more likely to postulate a large syntactic boundary at the location of a large prosodic boundary.

The present study is primarily concerned with Assumption (2a). Implicit in this assumption that silently generated implicit prosody resembles overtly produced explicit prosody is the idea that the latter can be used to study the former. However, as described below, previous work has challenged the validity of this assumption. The basic goal of the present study was to further explore implicit and explicit prosody's similarity, and to do so using methods that (a) separate questions about implicit prosodic form from questions about sentence processing patterns, and (b) take into account variation across individuals. To this end, the rest of the paper proceeds as follows. First, Section 1.2 describes work, carried out in the context of the IPH, that has led to doubts regarding Assumption (2a). Section 1.2 discusses evidence that Assumption (2a) might not be too far off the mark if two factors related to methodology are taken into account: (a) how explicit prosody is elicited, and (b) individual differences across speakers. Sections 1.3 and 2 then present an experiment testing for links between implicit and explicit prosody with these two factors in mind. We asked two basic questions. First, we asked whether individual differences in working memory capacity are systematically related to individual differences in explicit prosodic phrase length (observed in a production task), since working memory has been associated with implicit prosodic phrase length in previous work. Second, we asked whether individual differences in explicit phrase length observed in the production task were systematically related to individual differences in implicit prosodic phrase length, observed in a separate silent-reading task completed by the same speakers. A discussion of the results and concluding remarks follow in Sections 3 and 4.

1.2 *Implicit Prosody, Explicit Prosody, and Attachment Preference*

Although interest in the "internal voice" has a long track record in language research, the IPH ushered in a greater emphasis on the role that its prosodic aspects have on sentence processing (see also Bader, 1998). Work in this area has since expanded broadly, in terms of both the prosodic features investigated (e.g., phrasing, accentuation, lexical stress) and the psycholinguistic phenomena involved (e.g., syntactic parsing, word segmentation, lexical access).¹ It is therefore useful to

first consider Fodor's original motivation for the IPH, which was to account for patterns of ambiguity resolution involving sentences like 3:

- (3) *The doctor met the son of the colonel who tragically died.*

Such sentences feature ambiguity in the attachment of a relative clause (RC) to one of two preceding noun phrases (NP1 or NP2). If the RC (*who tragically died*) attaches low in the structure, it is interpreted as modifying NP2 (*the colonel*). If, on the other hand, the RC attaches high in the structure (to the entire complex noun phrase, *the son of the colonel*), it is interpreted as modifying NP1 (*the son*). As prosody was known to influence syntactic parsing for explicitly produced sentences (e.g., Maynell, 1999; Pynte & Prieur, 1996; see also Speer et al., 1996), Fodor proposed that prosody may be influencing the parsing of silently read sentences as well. The more specific claim in Fodor (2002) was that high attachment interpretations occur when the reader's inner voice inserts a prosodic boundary between NP2 and the RC: (NP1 NP2)/(RC). Low attachment interpretations occur when the reader instead inserts such a silent/implicit boundary between NP1 and NP2: (NP1)/(NP2 RC). Put simply, the RC modifies the constituent that it is grouped prosodically closest to, even when the prosody is only internally generated.

Two findings about RC attachment preferences serve as particularly strong support for the idea that implicit prosody influences syntactic parsing as per the IPH. The first is that readers' preference for high attachment increases when rhythmic pressures push for prosodic segmentation of the sentence, such as they do when an RC is very long (in phonological material). For example, it was known at the time (e.g., Fernández, 2003; Fernández & Bradley, 1999) and has since been well-replicated (e.g., Hemforth et al., 2015; Jun & Bishop, 2015a), that readers are more likely to report high attachment interpretations for sentences like (4a) than they are for variants like (4b) (from Rayner et al., 2012):

- (4) a. The doctor met the son of the colonel who tragically died of a stroke.
 b. The doctor met the son of the colonel who died.

The difference in attachment preference between (4a) and (4b) is thought to reflect the basic rhythmic constraints that are also at play in spoken prosody. While it is a matter of ongoing debate how strongly they are weighted relative to other constraints (Breen et al., 2011; Watson & Gibson, 2004), other things being equal, speakers prefer prosodic groupings that are neither too long nor too short (Krivokapić, 2007a; Nespore & Vogel 1986; Selkirk, 2000) and that are of roughly equal length (e.g., Cooper & Paccia-Cooper, 1980; Gee & Grosjean, 1983). Both of these rhythmic requirements push for dividing (4a) into two phrases, increasing the probability that a prosodic boundary will occur directly before the RC. This contrasts with the sentence in (4b), which may be fit comfortably into a single prosodic phrase, or perhaps two equal phrases, both of which group NP2 prosodically with the RC.

The second finding is that attachment preference is also systematically related to a property of readers rather than sentences—and in a way that is arguably related to the length effect just described. In a large individual differences study, Swets et al. (2007) found that whether or not native English-speaking readers report a high attachment interpretation for the relevant sentences is inversely related to their working memory capacity, as estimated by measures such as the reading span task (Daneman & Carpenter, 1980).² That lower working memory is associated with a preference for distal rather than local attachment is at first highly counterintuitive; if attachment to a local/more recent constituent is less taxing on processing resources (e.g., Frazier, 1979; Gibson, 1998; Kimball, 1973), we would expect individuals with lower working memory capacity to

instead prefer low attachment. Swets and colleagues argued this pattern could be explained if we assume two basic things: (a) that working memory capacity is positively correlated with the amount of material readers can fit into a “processing chunk,” and (b) that these processing chunks are prosodic in nature (albeit implicit prosody). These two assumptions then explain the counterintuitive pattern as follows. First, individuals with lower working memory, who must segment a string into smaller implicit prosodic chunks, are more likely to insert an implicit prosodic boundary before the RC. Then, as per the IPH, this implicit boundary increases the probability of a high attachment interpretation. The effect of working memory capacity can therefore be seen as related to the length effect on attachment described above; the only additional assumptions required are that what constitutes “too long for one phrase” varies from reader to reader, and that some of this variation is related to working memory resources. Given that a relationship between working memory capacity and the size of processing units has been independently demonstrated (e.g., Swets et al., 2014), the results of Swets et al. (2007) provide indirect but strong evidence for the psychological reality of implicit prosody. One of the questions we will ask in the present study is whether an analogous relationship exists between working memory capacity and explicit prosodic phrase length, which would seem to be predicted if implicit prosody and explicit prosody are similar in form.

Returning to Fodor’s original proposal, recall that her primary interest was in whether implicit prosody could help explain cross-language differences in attachment preference the way they could explain the within-language patterns just described (Carreiras & Clifton, 1993; Cuetos & Mitchell, 1988). It is in this line of research where Assumption (2a) has encountered the most difficulty. The logic in the relevant studies is as follows: if speakers of languages like Korean, Japanese and Spanish show an overall preference for high attachment (Cuetos & Mitchell, 1988; Jun & Kim, 2004; Kamide & Mitchell, 1997), it is because their default prosody phrases the RC separately from the two NPs. If speakers of languages like English and Romanian show a preference for low attachment (Ehrlich et al., 1999), it is because their default prosody phrases the RC together with NP2.³ Fodor’s Assumption (2a) is then called upon for indirect support: if these attachment patterns reflect language-specific prosodifications, we expect to see similar prosodifications of equivalent sentences that are explicitly produced. Assumption (2a) is thus a linking hypothesis of sorts, and crucially requires evidence from the analysis of speakers’ productions to confirm.

Unfortunately, confirmation has not been particularly strong. While groups of speakers of Korean (Jun 2003a; Jun & Kim, 2004), Japanese (Jun & Koike 2003) and Spanish (Bergmann et al., 2008), all high attachment languages, have been shown to preferentially produce the predicted boundary between the RC and the two NPs, native English speakers also produce this pattern (Bergmann et al., 2008; Jun, 2010; see also Jun & Shilman, 2008), despite English’s status as a low attachment language. While it should be noted that the explicit prosody of ambiguous RC sentences in low attachment languages other than English has been less well studied, the results so far are nonetheless discouraging. Jun (2010), Bergmann et al. (2008) and Breen (2014), for example, all conclude that the findings from explicit prosody for English seem to indicate that either implicit prosodic boundaries do not have the straightforward influence on attachment assumed by Fodor in (1), or it is too difficult to see, because implicit prosody and explicit prosody are not as equivalent as stated in Assumption (2a).

1.3 How Dissimilar are Implicit and Explicit Prosody?

While a growing body of work indicates that the IPH underestimated the complexity of the relationship between prosody and off-line attachment preference,⁴ evidence based on a number of psycholinguistic and neurophysiological methods suggests it was not fundamentally off the mark (e.g., Hirotani et al., 2006; Jun & Bishop, 2015a; Jun & Bishop, 2015b; Schremm et al., 2015;

Steinhauer, 2003; Steinhauer & Friederici, 2001; Swets et al., 2007). Rather than a fundamental error in diagnosing the relationship between implicit prosody and syntactic parsing, the problem instead lies in the diagnosing of implicit prosody itself—and in particular, assuming its similarity to the explicit prosody obtained from reading aloud. Said otherwise, the problem may fundamentally be a defect in Assumption (2a).

Further complicating matters is the possibility that rather than a problem with reading aloud *per se*, the disconnect may have more to do with the details of how explicit prosody has been elicited in previous studies, including ones like Jun (2010). As Jun (2010) notes in her discussion of the matter, there is considerable evidence that reading aloud is more closely related to phonological processing, while silent reading is more closely associated with semantic processing (Bookheimer et al., 1995). Prosody from read-aloud speech is therefore more likely to lack influences from syntactic and semantic structure than is implicit prosody generated during silent reading (or even explicit prosody produced in conversational speech). For example, when reading individual sentences aloud in a laboratory, and especially when sentences are long or complex, speakers likely prioritize fluency and rhythm over the encoding of syntactic structure and semantic meaning (e.g., Rasinski, 2006). This may be exacerbated for individual sentences read in isolation, given that meaning is less effectively processed in the absence of a larger semantic/pragmatic context (e.g., Turnbull et al., 2017). Moreover, structure and meaning may also be weakly encoded in prosody if the speaker prosodifies the sentence before fully comprehending it (Jun, 2010), is insufficiently aware of the presence of ambiguity (Hwang & Schafer, 2009; Snedeker & Trueswell, 2003), or lacks the communicative motivation to explicitly reduce ambiguity (Breen et al., 2010). The corollary of this, however, is that if readers have time to comprehend the material (and some context is provided to help them do this), even relatively lab-speech-style reading tasks may be informative as to readers' implicit prosodic representations (see Hwang & Schafer, 2009, for additional discussion on this point).

These conclusions find some support in a study demonstrating a relation between implicit and explicit accentuation (i.e., prominence patterns rather than phrasing patterns) using speech from reading aloud. In two experiments, Speer and Foltz (2015) investigated readers' generation of implicit intonational pitch accents for sentences in different focus contexts. Participants in their study were presented with sentence pairs that either placed (5a) or did not place (5b) contrastive focus on the subject of the second sentence in a pair (the "test sentence," here in bold):

- (5) a. Jacquelyn didn't pass the test. **Belinda passed the test.**
b. Belinda didn't fail the test. **Belinda passed the test.**

Described in terms of ToBI (Tones and Break Indices) annotations for English (Beckman & Ayers Elam, 1997; Beckman & Hirschberg, 1994; Silverman et al., 1992), the canonical intonational structure of the test sentence in (5a) would include a prominent L+H* pitch accent on the contrastively focused *Belinda*, but no pitch accent on the given information following it (Büring, 2016). In (5b), where the verb is contrastively focused and the other information given, a L+H* is expected on *passed*, and the other information of lower prominence/unaccented. Speer and Foltz reasoned that if readers were implicitly generating these prosodic patterns, it should be detectable via priming effects on the recognition of explicit (i.e., auditorily presented) prosodic forms. To test this, participants in their study first silently read sentences like those in (5) and were then presented with an auditory target—an auditory version of the sentence's subject, which either matched or did not match the "canonical" prosody. Participants were then asked to decide whether or not the auditory target was the same word that began the test sentence they had just read. The prediction was that, if participants were implicitly generating the canonical intonation, there should be an advantage for responses to auditory targets that match that canonical intonation. After reading (5a), this should mean an advantage for

Belinda pronounced with a L+H*; after reading (5b), it should mean an advantage for *Belinda* pronounced as unaccented. Surprisingly, the results of this first experiment revealed no significant overall effect that indicated priming; responses to L+H* auditory targets were not faster following contrastively focused primes for their participant group as a whole.

Interestingly, however, this group-level result did not hold true for all participants, as some showed fairly robust priming for implicit-explicit matches. Having identified this variation across participants, the authors hypothesized that the lack of overall priming effects may have occurred not because readers failed to generate implicit accentuation, but because they did not all generate the *same* implicit accentuation. If, for example, some participants were generating the less-prominent H* pitch accent for contrastively focused *Belinda* rather than an L+H* (which is in fact not uncommon in explicit speech; e.g., Calhoun, 2012; Ito & Speer, 2006), then priming effects would not be expected. To test this, they repeated their first experiment, but included an additional reading-aloud task that was intended to identify individual differences in (explicit) speech style that might correlate with priming effects. Before engaging in the priming task, participants in their second experiment first read a brief (≈ 100 words) passage aloud. This passage was used to distinguish individuals who reliably used L+H* to mark contrastive focus from individuals who used it less reliably (in favor of a less-prominent H* accent). In fact, the results of this second experiment showed that participants who produced more L+H* pitch accents in the read-aloud passage were the same participants who showed a priming effect for L+H*/contrastive focus matches; participants who tended to produce more H* pitch accents, on the other hand, were the participants who failed to show the priming effect.

Particularly relevant to the issue of phrasing patterns, Speer and Foltz highlight that sentence-level prosody is rife with optionality, even in terms of categorical contrasts. Thus, unlike in the study of word-level phenomena (see Breen, 2014 for several examples), it may simply not be possible to study implicit sentence-level contrasts like accentuation and phrasing without considering cross-speaker (and cross-reader) variation. Notably, Speer and Foltz's reading-aloud task, though far from eliciting natural or spontaneous speech, made use of a passage that was longer and had richer context than individual sentences in isolation do. This likely helped speakers process its meaning, leading to better prosody-to-meaning mapping. However, rather than attempting to relate explicit prosody for isolated sentences to sentence comprehension patterns, the read-aloud passages were used to form a picture of overall speech style for individual speakers. That is, based on a larger sample of speech, Speer and Foltz determined what was a *probable* implicit prosodic pattern for *particular individuals*.

1.4 Present Study: Individual Differences and Prosodic Phrasing

Given that most of the work on implicit prosody has been carried out in the context of the IPH, it has been difficult to separate questions about implicit prosodic form from questions about syntactic ambiguity resolution. As our interests center around the former rather than the latter, the present study attempted to investigate implicit prosody without considering attachment preference. More specifically, the overarching goal was to probe for similarities between the implicit prosody generated in silent reading and the explicit prosody produced in reading aloud. To this end, the experiment below asked two questions:

- Q1. Do individual differences in working memory capacity predict individual differences in explicit prosodic phrasing?

This first question derives largely from the findings in Swets et al. (2007) and Speer and Foltz (2015), both of which demonstrate the relevance of individual differences. As described above, Swets and colleagues make the claim that working memory capacity (henceforth WMC) influences the size of

the prosodic chunks that readers implicitly generate when reading text silently. This offers us an opportunity to link implicit and explicit prosody because it leads to the following prediction (which has not, to our knowledge, been tested): if implicit and explicit prosody are similar in form, we expect WMC to have the same basic relationship with explicit phrase length as it does with implicit phrase length.⁵ Much like in Speer and Foltz (2015), the explicit prosody we tested in the present study was elicited in a simple reading-aloud task, using a narrative-like passage. Such materials may encourage better processing of meaning by providing more depth and coherence than does a list of isolated sentences. To encourage speakers to produce the kind of prosody consistent with their interpretation of meaning in the passage, speakers were given time to skim the entire passage, and also practiced reading it aloud before recording the sample to be analyzed. Speech produced in this reading-aloud production task was then annotated for its phrasal structure, phonologically defined, and the extent to which WMC predicted variation in prosodic phrasing (i.e., phrase length) across speakers was modeled statistically.

Q2. Do individual differences in explicit prosodic phrasing predict individual differences in self-identified implicit prosodic phrasing?

The second question addresses the relation between explicit and implicit prosody in a different way. The basic idea was that while implicit prosody cannot be observed directly, readers may be able to provide information about its form via self-reports, which can then be compared against their more directly observable explicit prosody. Therefore, in a second task, the same participants who served as speakers in the reading-aloud task just described were also asked to read a second passage silently, and to indicate where in this passage they perceived their “inner voice” to be inserting prosodic juncture.

While the task may be challenging, there is reason to think it can provide some insight into the inner voice. Recent work on the use of “crowdsourced” prosodic annotation (using “Rapid Prosody Transcription”; Cole, Mo, & Baek, 2010; Cole, Mo, & Hasegawa-Johnson, 2010; Cole et al., 2017) has demonstrated that linguistically untrained listeners have some aptitude for identifying auditorily presented prosodic events, and it is not clear a priori that the perception of internal events, like those occurring in inner speech, is completely different from the perception of external events. Moreover, the psychological reality of inner speech, as discussed in Fodor (2002), has never been seriously in doubt, nor has the ability of most people to attend to it to some degree (see also Rayner et al., 2012; and Slowiczek & Clifton, 1980). And while we might expect considerable variation in readers’ ability to identify implicit prosodic events, we also expect variation in listeners’ identification of external, auditorily presented ones (Bishop et al., 2020; Roy et al., 2017). Moreover, the goal here was not to use their responses as a stand-alone picture of their implicit prosodic representations. Instead, the purpose was to *relate* these self-identified implicit prosodic events to the directly observable explicit prosody produced by the same individuals. The prediction was that individuals who report parsing written text into smaller prosodic units implicitly should also tend to parse read-aloud sentences into smaller units in explicitly produced speech. To the extent that they do, this would provide new evidence for the similarity between explicit and implicit prosodic form, using cross-speaker/reader variation to do so.

2 Experiment

2.1 Methods

2.1.1 Stimuli. Materials were chosen for two reading tasks: a reading-aloud task (to assess explicit prosodic phrasing) and a silent-reading/implicit-boundary identification task (to assess implicit prosodic phrasing). In both cases, passages of moderate length were used (taken from popular

prose writing in Alda, 2006), chosen for their simple but engaging narrative and large proportion of short words (approximately 75% of the words in each passage were monosyllabic). A high proportion of short words was particularly desirable because it allows for greater variation in phrasing across speakers/readers; given a set amount of text, shorter words means more word boundaries, and word boundaries are the potential locations of phrase boundaries. The passage to be read aloud by speakers was 156 words, and the passage to be read silently was somewhat longer at 255 words. (It was assumed that a larger sample would be needed to detect individual differences in the silent-reading task, given the potential difficulty participants may have in identifying their own implicit prosodic boundaries). A sample of the passage used in the reading-aloud task is shown in (6); the passages used in both tasks are shown in their entirety in the Appendix.

- (6) *Our patch of tangled yard was an exotic foreign country. I had spent so much of my life in dark theaters and dim hotel rooms, where the only thing green was the peeling paint on the walls, that this seemed perfectly natural to me. This was where I had my first bite of mud pie; where I set up a card table and mixed household chemicals, toothpaste, and my mother's face powder, doing what I called "experiments". . .*

2.1.2 Participants. Participants were 100 native English speakers (42 male, 58 female), mostly undergraduate students living in one of the boroughs of New York City. All participants self-identified as free of any history of hearing, communication or learning disorders, and all were compensated for their participation.

2.1.3 Procedure. All participants engaged in the following tasks (and in the following order): completion of the Autism Spectrum Quotient questionnaire ("AQ"; Baron-Cohen et al., 2001); completion of an automated version of the reading span task (Daneman & Carpenter, 1980); a reading-aloud production task; a silent reading and boundary identification task; and a computer-based version of the Stroop Test (Golden, 1978; Stroop, 1935). The procedure for the two reading tasks, intended to elicit measures of explicit and implicit prosody, respectively, are described below. The AQ and Stroop measures (collected as part of a larger study of individual differences in speech) were unrelated to the present investigation and are not discussed further. The reading span task was carried out using an E-Prime implementation created by Oswald et al., (2015). In brief, participants were presented, on each trial, with a string of alphabetic letters to hold in memory while performing another task, namely the reading of a sentence of 10–15 words. Participants were then required to judge whether or not the sentence made sense before being asked to recall the original letter string (in the correct order). Readings spans for each participant were estimated based on the "partial-score" method in Conway et al., (2005), reflecting the total number of trials with accurate recall of strings (rather than the "absolute" scoring based on a limited number of sets of trials for which perfect accuracy was achieved), which allows for maximal variation across participants to emerge.

For the production task, participants read aloud the 156-word passage while seated in a sound-attenuated booth. Participants were asked to first skim the passage silently to familiarize themselves with its content; they were then to read the passage aloud into a Shure SM10A head-worn unidirectional dynamic microphone (recorded digitally at 44.1kHz), twice, at a comfortable rate and in a style that was (in their own assessment) natural and appropriate for the passage's content. The first recorded reading was intended as practice, in order to increase fluency. The second recorded reading was saved as a WAV file and retained for later analysis of its phrasal structure, phonologically defined within the Autosegmental-Metrical framework (Beckman & Pierrehumbert, 1986; Ladd, 2008; Pierrehumbert, 1980), and more specifically using the ToBI conventions for Mainstream American English (Beckman & Ayers Elam, 1997).

For the silent-reading/implicit-boundary identification task, participants first underwent training in the identification of prosodic boundaries. (It was determined that no participant had prior experience/training in prosodic theory or prosodic transcription). The first part of this training involved listening for explicit prosodic juncture in auditorily presented running speech; participants were presented with a short practice recording (55-words in length, taken from one of former President Barack Obama's Weekly Addresses)⁶ for which they carried out a Rapid Prosody Transcription task (e.g., Cole, Mo, & Baek, 2010), listening for instances of explicit prosodic juncture. The instructions given to participants were to listen for places where the speaker produced pauses or very subtle "pause-like" separation between words. It was explained that this "separation" was not always an actual silent pause in speech, but rather anything done with the voice to make the words "sound separated from each other, as if belonging to different speech chunks." This was described as sometimes being the result of speakers slowing down their speech rate, or a change in pitch, at the end of a "chunk." Participants listened to the practice recording three times, and marked where they heard such juncture. In terms of ToBI categories, the recording used contained instances of both large Intonational phrase (IP) boundaries and more subtle intermediate phrase (ip) boundaries. The short 55-word passage and locations of such boundaries (as labeled by a trained ToBI transcriber) are shown in (7):

- (7) *This week I spent some time in Colorado // and Texas, // talking with people about what's going on in their lives. // One of them was Elizabeth Cooper, // who will be a college junior this fall. // She wrote to tell me something I hear often: // how hard it is / for middle class families / like hers // to afford college.*

After hearing the recording the third time, the experimenter made sure that the subject had identified all of the prominent IPs and more subtle ips (labeled in (7) as "//" and "/", respectively, although the participants only used a single slash, and did not differentiate juncture of different degrees above the word level). Participants who did not identify all of the labeled boundaries were permitted to listen to them again until they reported being able to identify the difference between these more subtle breaks and the absence of one. Generally, listeners did not have trouble identifying these particular instances of phrase boundaries on their own, and seemed able to understand the task. Importantly, what was highlighted to them was that such separation between words often occurs in the absence of orthographical punctuation like a period or comma, and that the goal was to identify all cases of juncture, regardless of punctuation in the text.

The second part of the training involved a practice run for identifying implicit boundaries in their own inner speech during silent reading of a passage of similar length (62 words) and content. Participants were introduced to the idea of the "voice in your head" that people use when reading silently. This idea was highly intuitive to nearly all the participants in the study. They were asked to read the 62-word practice passage silently at their own pace, and to try to mark the ends of "chunks" in their inner speech, just as they had done with the external voice in the first part of the training. It was explained that there were no "right" or "wrong" responses, and that responses were expected to be different for different readers. Participants carried out the task at their own pace (and thus re-reading was assumed to be taking place); responses on these practice materials were discarded and thus not available for later analysis.

Following this training session, participants were then presented with the 255-word passage (again, shown in the Appendix), and asked to carry out for this longer passage the same silent-reading/juncture-perception task they had just practiced. Participants' identification of prosodic juncture in this longer passage was retained for later analysis.

2.2 Analysis

2.2.1 Production task: identification of explicit boundaries. The purpose of the production task was to identify variation in how speakers group the passage into prosodic phrases, and in particular, the average length of their prosodic phrases. As all participants in the experiment read the same passage, it can be assumed that (a) differences in speakers' prosodic phrases were for the most part not due to the lexical/syntactic/semantic content of the passage, and (b) the length speakers' phrases would be, on average, inversely related to the number of prosodic boundaries they produced in the passage. It is important to emphasize that the prosodic junctures of interest here were those that were intended by the participants—that is, the output of fluent speech. To this end, it was necessary to identify both fluent and disfluent instances of juncture so that the latter could be removed from the analysis.

Annotation of the speakers' productions was therefore carried out as follows. Due to the considerable size of the speech corpus collected (the 100 participants produced a total of approximately 75 minutes of speech), a modified version of the ToBI transcription process was used, designed to spread out the labor-intensive task of phonological annotation among multiple researchers. First, one senior and one junior researcher trained in ToBI transcription (working together as a pair) identified in each participant's recording the locations of (a) all disfluencies and (b) all potential fluent prosodic boundaries. "Disfluencies" were identified as per the ToBI guidelines (Beckman & Ayers Elam, 1997), but were additionally categorized according to commonly recognized types (e.g., Daneman, 1991; Engelhardt et al., 2013; Lickley, 2015). In particular, a disfluency was identified as either a *hesitation* (an accidental prolongation), a *repetition* (an apparently unintended repetition of a word or string of words), a *repair* (halting the production of a word after starting it, and then starting over with another word), or a *filled pause* (the insertion of syllables like "uh," "um," and "er"). "Potential fluent prosodic boundaries" were identified as the locations of any non-disfluent perceived juncture greater than that marking an ordinary word boundary (i.e., anything corresponding to a 2, 3-, 3, 4-, 4, or any uncertainties (e.g., a 2? or 3?) in the ToBI annotation conventions).

These first-pass "potential" boundary identifications (which resulted in a total of 4,467 potential prosodic boundaries for the group) were intended to be liberal; they would then serve as the input to a more conservative second-pass ToBI transcription carried out by two additional trained ToBI labelers who made final decisions about them. For each of the potential boundaries, the two labelers made decisions regarding two things: the break index and any edge-marking tones. The break index was identified via forced choice as either *no prosodic break* (a ToBI word-level boundary 0, 1, 1-, 2, or "?-"), an *ip-level boundary* (a ToBI 3- or 3, or 3?) or an *IP-level boundary* (4-, 4, or 4?). Edge-marking tones were identified as either a T- tone (corresponding to the edge of an ip) and a T% tone (corresponding to the edge of an IP). Boundary tone distinctions themselves had no part in the analysis; including them in the transcription was primarily intended to ensure that the labelers were making use of tonal information in their decisions about boundaries).

Agreement levels between the two labelers were generally consistent with what has been reported elsewhere for the ToBI conventions (Breen et al., 2012; Pitrelli et al., 1994; Syrdal & McGory, 2000; Yoon et al., 2004).⁷ Agreement between labelers for the presence of a prosodic break above the word level was 92.5% ($\kappa = .76$), and 100% of the disagreements concerned the presence of an ip boundary versus the absence of a boundary (i.e., there were, unsurprisingly, no cases where one labeler identified a word-level boundary and the other an IP-level boundary). Where both labelers agreed a boundary above the word level was present, agreement as to its size (i.e., an ip boundary vs. an IP boundary) was 94.5% ($\kappa = .88$). Because the analysis required definitive decisions about the locations and sizes of boundaries for all speakers, and because we wished to analyze only those

boundaries about which we had the most certainty, the relatively small proportion of disagreements were settled in the direction of the labeler who marked a smaller degree of juncture. That is, for disagreements where one labeler assumed a word-level boundary and the other an ip-level boundary, the word-level boundary was assumed; in cases where one labeler marked an ip-level boundary and the other an IP-level boundary, the ip-level boundary was assumed. As a result of this annotation method, of the 4,467 potential boundaries initially identified in the first pass, a total of 3,427 actual prosodic boundaries were identified for final analysis (61.4% of which were IP boundaries).

Finally, the ToBI-defined prosodic phrase boundaries were used to quantify participants' individual tendencies regarding phrase length. We took the number of fluent phrase boundaries as a measure of average phrase length; as all speakers read the same passage with the same lexical material, the number of boundaries is directly indicative of (and inversely related to) the length of those phrases. These phrase boundary counts either disregarded the prosodic hierarchy (i.e., we considered the binary presence versus absence of phrase breaks, ignoring whether a phrase boundary marked the edge of an intermediate phrase or an IP) or considered lower-level (intermediate) and higher-level (Intonational) phrase boundaries separately. The reason for considering these different kinds of prosodic grouping was due to the fact that it was not clear a priori which level of prosodic structure WMC—if it relates to speakers' phrasings at all—it would be most closely associated with. For example, lower WMC may be associated with a tendency to group utterances into a greater number of shorter IPs (which would be inferable from the number of intonation phrases they produced). Alternatively, it may be associated with a tendency to group the material *within* longer IPs into a series of shorter intermediate phrases (detectable in the number of intermediate phrases contained within the speakers' IPs). Finally, it is possible that lower WMC might be associated with shorter phrases at both levels of structure.⁸

These different methods of counting speakers' use of phrase boundaries served as the measures for analysis, and are summarized in Table 1. When considering these counts, it is important to note that, as intermediate phrases and Intonational phrases are hierarchically organized (with the former dominated by the latter), the location of every IP boundary is also the location of an ip boundary, but not every ip boundary corresponds to an IP boundary. Therefore, the "number of ips" measure used here refers to only those ips that did *not* coincide with the right edge of an IP (that is, it includes only non-IP-final ips). By the same token, the "total number of boundaries" measure is equivalent to all ips, since every boundary is necessarily an ip. Finally, as it is well-known that speech rate is inversely and sometimes very strongly correlated with boundary placement (e.g., Fougeron & Jun, 1998; Gee & Grosjean, 1983; Jun, 1993; Jun, 2003b), we also estimated each participant's overall speech rate, defined as their average syllable duration, calculated over all ips in the passage (and excluding any silent pauses that could not be better explained by voiceless stop closure).

2.2.2 Silent-reading task: identification of implicit boundaries. The quantification of participants' implicit boundaries in the silent-reading task was considerably simpler, and was based on the junctural annotations that the participants themselves provided on the printed transcripts of the second passage. Again, since all participants read the same passage, a greater number of boundaries marked indicates shorter implicit prosodic phrases on average. Clearly here, however, there can be no distinction between ip- and IP-level boundaries, since readers were only instructed to make binary judgments about the presence of juncture between words in the printed passage, and had no training or other basis for making finer distinctions.

2.3 Results

2.3.1 Production task: explicit prosodic phrasing and WMC. We begin first with a brief description of disfluencies produced by participants. Disfluencies have sometimes been found to be inversely

Table 1. Measures of phrase length used for analysis.

Number of ips produced	Total number of fluent intermediate phrase boundaries produced in the passage (excluding those that coincided with an IP boundary)
Number of IPs produced	Total number of fluent Intonational phrase boundaries produced in the passage
Total number of Boundaries	Total number of fluent prosodic boundaries (of any phrase level, intermediate or Intonational) produced in the passage

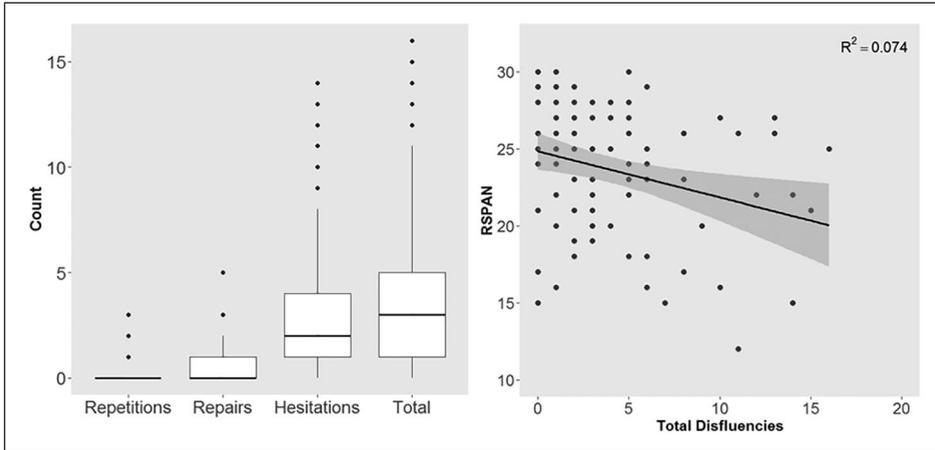


Figure 1. (Left) Box plot summarizing the prevalence of different disfluency types produced by speakers. (Right) Scatterplot showing the relationship between WMC (as estimated by reading spans) and disfluencies for the group of speakers. (Higher RSPANs indicate higher WMC).

related to WMC (e.g., Daneman, 1991) and for this reason it was important to identify and exclude them from the main analysis, which was concerned with WMC's possible relation to fluent, intended prosodic groupings. A breakdown by disfluency type is shown in Figure 1 (left panel).⁹ As a group, speakers produced an average of four disfluencies for the 156-word passage (most of which were hesitations/prolongations), approximately half of what some studies have reported for other materials and speech styles (Bortfeld et al., 2001; Fox Tree, 1995; Shriberg, 1994). We also briefly explored whether this sample of disfluencies was related to WMC. As can be seen in Figure 1 (right panel), an inverse relationship is apparent, with reading spans accounting for approximately 7.5% of the variance according to a simple R^2 . A mixed-effects logistic regression model was then used to test the significance of this relationship; the model attempted to predict the categorical presence of a disfluency (of any type) in speakers' readings of the passage as a function of fixed effects (participants' reading spans and speech rate, centered on their means) and random intercepts for participant and the word's linear order in the passage (a measure of how far along in the passage the speaker was). This revealed significant simple effects for both fixed-effects factors; lower reading spans (*estimate* = -0.055, *ste* = .02626, *z* = -2.08, *p* < .05) and slower speech rate (*estimate* = 14.00, *ste* = 5.0157, *z* = 2.79, *p* < .01) were both associated with a higher probability of disfluency. Thus, as has been reported previously, speakers with lower readings spans (indicating lower WMC) tended to produce speech more frequently interrupted by disfluencies. However, it is important to note that the overall number of disfluencies in this sample was, by design, quite

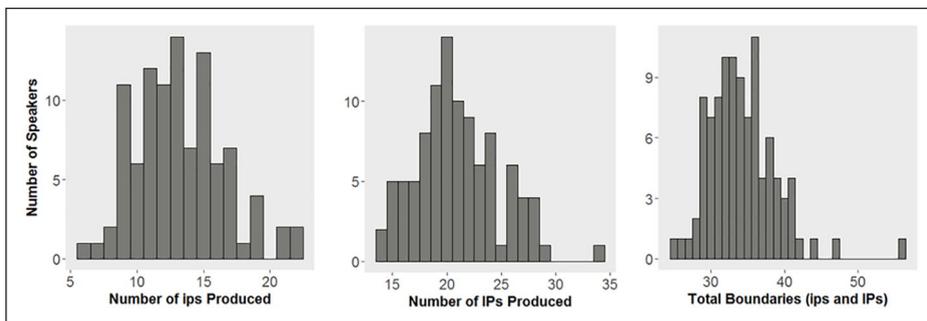


Figure 2. Histograms plotting the distribution of participants' explicit prosodic groupings, in terms of intermediate phrases (ips), Intonational Phrases (IPs) or total number of boundaries, ignoring phrase level.

small. Furthermore, the primary reason for identifying these disfluencies was so that they (and any relationship they have to WMC) could be removed from the analysis of fluent prosodic phrasings, to which we now turn.

In considering whether individual differences in WMC predict cross-speaker variation in phrase length, it is clearly important that considerable cross-speaker variation actually exists. In fact, Figure 2, which shows how speakers were distributed in terms of the number of phrase breaks they produced, indicates substantial variation in this group. That is, even when reading the same, relatively brief scripted passage, there was a considerable lack of uniformity in terms of prosodic phrasing. This indicates that there must be variation that the grammar (i.e., syntactic, semantic, and rhythmic constraints) and the orthography of the passage are unable to account for. The purpose of the rest of the analysis was therefore to determine whether differences in speakers' WMC accounted for some of this variance—and, if so, if it was tied to a particular level of the prosodic hierarchy.¹⁰

Figure 3 plots participants' phrasing as a function of reading spans. As can be seen, a negative correlation is present between readings spans and each measures of phrasing considered, indicating that lower WMC was associated with a greater number of prosodic boundaries (and thus shorter phrases) in the passage. Notably, the correlation is of similarly modest strength when considering intermediate phrasing and Intonational phrasing separately; the correlation is strongest (accounting for as much as approximately 20% of the variance) when collapsing for phrase types.¹¹ This would seem to suggest that WMC is predictive of the amount of material speakers parse into prosodic phrases of some type—but not which type (ip or IP) it is.

To test the statistical significance of these correlations in a more rigorous way, mixed-effects logistic regression was used. First, a set of models was constructed to predict either the categorical presence of an ip boundary, an IP boundary, or a boundary of any size (i.e., ignoring distinctions in phrase level). In these models, the fixed effects included the categorical predictor “punctuation” (the presence or absence of punctuation following a word in the text) and the continuous predictors “reading span” and “mean syllable duration.” “Mean syllable duration,” the measure of speech rate used here, is especially important. As discussed above, it is well-known that speakers tend to omit prosodic boundaries when speaking at faster rates, and preliminary inspection of the data revealed that mean syllable duration was also negatively correlated with reading span ($R^2 = .102$), indicating that speakers with lower WMC tended to read at a slower rate. This raises the possibility that the correlations between reading span and boundary placement described above could have been accounted for by speech rate, and so speech rate must be statistically controlled for. The random effects included those that contributed significantly to model fit as indicated by a log-likelihood

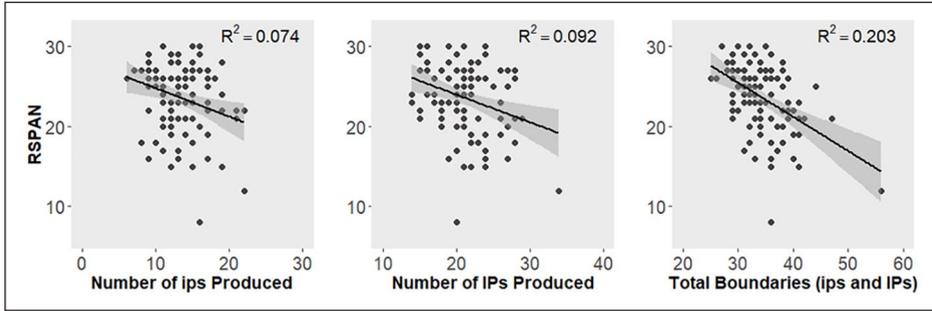


Figure 3. Scatterplots showing the relationship between participants’ WMC (as measured by reading spans) and the prevalence of prosodic boundaries (defined in terms of intermediate phrases (ips), Intonational Phrases (IPs), or total number of boundaries of any type). Higher RSPANs indicate higher WMC.

ratio test (Matuschek et al., 2017). These included, in addition to intercepts for “participant” and “word” (conceived of here as an “item”), a by-participant slope for “mean syllable duration.” Continuous variables were centered on their means.

The results of the logistic regression models are shown in Table 2. First, and of primary interest, reading span was a statistically significant predictor of participants’ prosodic phrasing only when collapsing for phrase level—indicating that WMC’s relationship was with phrase boundaries in general, not phrase boundaries marking a particular level in the prosodic hierarchy. Second, and unsurprisingly, the models confirmed that the presence of punctuation in the text predicted prosodic juncture. However, this was only significant when modeling IP boundaries or boundaries in general; this is expected, as participants likely produce large prosodic breaks (often followed by pauses) rather than more subtle ip boundaries at points of orthographic punctuation. Also unsurprising is the fact that longer mean syllable duration (i.e., slower speech rate) was significantly associated with the likelihood of prosodic boundaries. This was true for all measures of phrasing; this is consistent with previous work showing that faster speech rates can lead to prosodic restructuring such that, at faster speech rates, boundaries are demoted from a higher level to a lower level, and boundaries marking a lower level are sometimes deleted entirely (e.g., Fougeron & Jun, 1998).

To summarize, this part of the study sought to determine whether variation in the prevalence of prosodic boundaries (and thus average phrase length) in read-aloud speech was significantly related to speakers’ verbal WMC. In fact, such a relationship was observed, and was statistically significant; reading spans were negatively associated with the number of boundaries speakers produced (and thus positively associated with average phrase length). The results from this group of English speakers therefore indicate that WMC has the same basic relation to explicit prosodic phrasing that Swets et al. (2007) argued exists between WMC and implicit prosodic phrasing. This is broadly consistent with the idea that implicit prosody and explicit prosody are similar in form—and thus also lends some support to the validity of Assumption (2a) of the IPH.

2.3.2 Implicit boundary identification task: Implicit and explicit prosodic phrasing. Recall that the purpose of this analysis was primarily to relate speakers’ explicit prosodic phrasing patterns (in production) to their own implicit prosodic chunking patterns (in silent reading). As a first-pass assessment of how explicit and implicit prosodic phrasing behavior correlated, Figure 4 (top panel) plots the number of implicit prosodic boundaries participants identified in the silent-reading task with the number of explicit prosodic boundaries they produced in the reading-aloud production task. The figure also shows (bottom panel) the number of implicit breaks participants identified as a function of their

Table 2. Results of three mixed-effects logistic regression models predicting either the presence of intermediate phrase boundaries (ip), Intonational phrase boundaries (IP), or a phrase boundary of any type. (The intermediate phrase boundaries modeled were only those that did not correspond to an Intonational phrase boundary).

	Estimate	STE	z	p
ip Boundaries:				
Punctuation	-1.3078	.0101	-1.62	> .1
Mean syllable duration	.0448	.0205	2.18	< .05
RSPAN	-.0170	.0101	-1.69	< .1
IP Boundaries:				
Punctuation	13.0005	1.8444	7.05	< .001
Mean syllable duration	.0257	.0392	6.56	< .001
RSPAN	-.0227	.0200	-1.13	> .1
Any Boundary:				
Punctuation	9.7900	1.2878	7.60	< .001
Mean syllable duration	.0241	.0263	9.17	< .001
RSPAN	-.0407	.0136	-2.99	< .01

reading spans. Apparent in the figure is that participants' self-identified implicit prosodic boundaries were very weakly related to their reading span, although the direction of the correlation was nominally negative. The relationship between participants' implicit boundaries and explicit boundaries was also generally weak; essentially no correlation was found between implicit chunking and explicit ips, or between implicit chunking and the total number of explicit boundaries. The strongest correlation was between implicit boundaries and explicit IPs, a positive correlation, although explicit IPs accounted for only approximately 5% of the variance according to a simple R^2 .

To test whether participants' explicit IP boundaries were able to significantly predict their implicit prosodic boundaries in a more complete statistical model (and to confirm that explicit ip boundaries and total number of explicit boundaries did *not*), mixed-effects logistic regression models were fitted to the data in a manner similar to the analysis of explicit phrasing. Here, three models were constructed, each attempting to predict participants' self-identified implicit prosodic boundaries as a function of a different explicit prosodic variable (i.e., explicit ips, explicit IPs, or the total number of explicit boundaries a participant produced in the reading-aloud task). Punctuation (fixed effect) and reading spans (fixed effect and by-participant random slope) were again included in the models. (Reading spans were not included in the model that tested the "total boundaries" measure, as they did not contribute to model fit; Matuschek et al., 2017).

The results of modeling are shown in Table 3 (testing ips), Table 4 (testing IPs) and Table 5 (testing "total boundaries"). Unsurprisingly, in all models, overt punctuation in the passage's text strongly predicted readers' reporting of implicit juncture. Self-reports of implicit juncture were also significantly associated with speakers who produced more IP boundaries in the reading-aloud task, the only explicit prosodic measure that was relevant. Finally, participants with shorter reading spans were also numerically associated with fewer implicit prosodic breaks, but not significantly so.

To summarize, this part of the study explored the relation between explicit and implicit prosody using a novel self-report method of implicit prosodic chunking. While participants' explicit prosody was not, overall, a particularly robust predictor of their self-reported implicit prosody, some hints at a relationship between the two were detectable—and interpretable. The main finding was that participants who produced more IPs in the production task tended to report more implicit breaks in their silent reading, and this was statistically significant. One possibility is that this

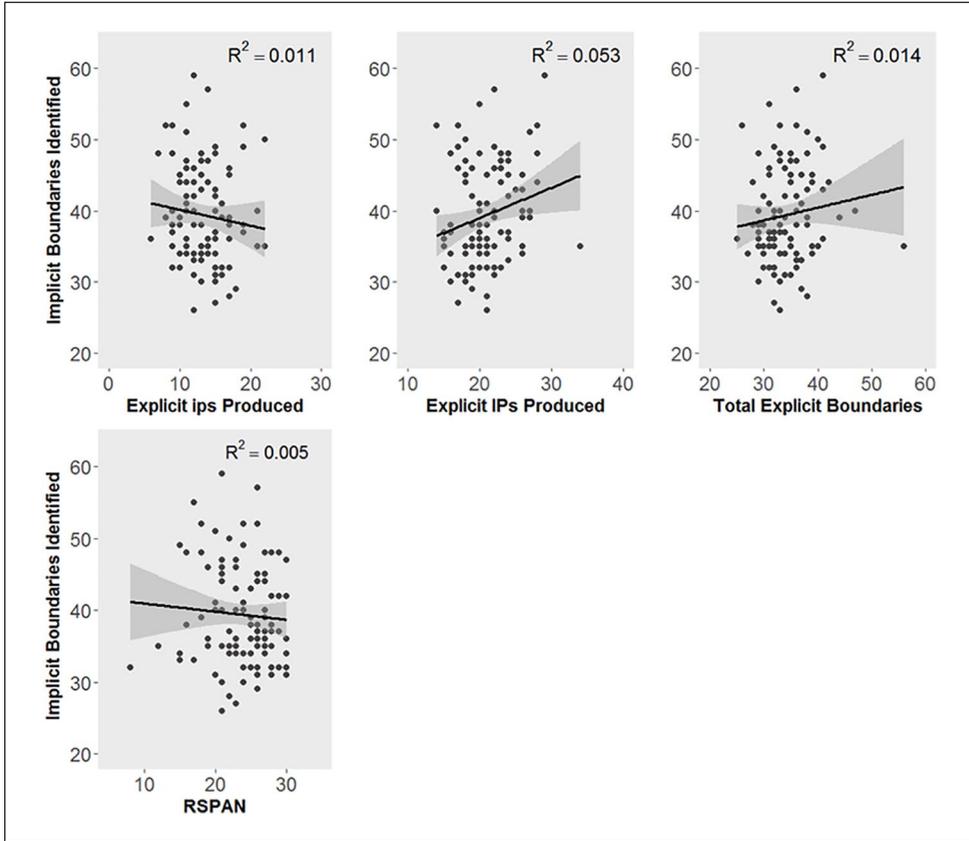


Figure 4. Scatterplots showing the relationship between speakers' implicit and explicit boundaries (top) and the relationship between their implicit boundaries and RSPAN scores (bottom).

reflects a dissimilarity between implicit and explicit prosody; perhaps, for example, implicit prosody has a simpler structure, one that consists of only higher-level (IP) groupings. Related to this, it is possible that implicit and explicit prosody are only faintly similar in form, but both related to a common, deeper representation of structure.

Yet another interpretation—one that seems particularly plausible—is that what we found here instead reflects the difficulty of the task employed. That is, participants may only have been proficient at identifying prominent, higher-level IP boundaries in their inner speech (and even then, likely only a subset of them), while missing most of the subtler ip boundaries. Thus, while the methodology used in this task may need to be modified in order to be sufficiently sensitive for wide use in investigating implicit prosody, it did seem to detect a relationship between explicit prosody and implicit prosody above chance level. Like the results from the first task, this provides some support for the claim that implicit and explicit prosody are similar in form—supporting Assumption (2a) of the IPH.

Table 3. Results of the mixed-effects logistic regression model used to predict participants' self-reported implicit prosodic boundaries as a function of the number of explicit *intermediate phrases* they produced.

	Estimate	STE	z	p
Punctuation	8.6468	.6022	14.35	< .001
RSPAN	-.0281	.0184	-1.52	> .1
Explicit ips	-.0311	.0210	-1.48	> .1

Table 4. Results of the mixed-effects logistic regression model used to predict participants' self-reported implicit prosodic boundaries as a function of the number of explicit *Intonational Phrases* they produced.

	Estimate	STE	z	p
Punctuation	8.631	.5984	14.42	< .001
RSPAN	-.0115	.0181	-0.63	> .1
Explicit IPs	.0459	.0202	2.27	< .05

Table 5. Results of the mixed-effects logistic regression model used to predict participants' self-reported implicit prosodic boundaries as a function of the *total number of explicit prosodic boundaries* they produced.

	Estimate	STE	z	p
Punctuation	8.633	.6007	14.37	< .001
Total Explicit Boundaries	.0158	.0157	0.99	> .1

3 Discussion

The study presented above was motivated by questions regarding the similarity between implicit prosody and explicit prosody, and thus the appropriateness of attempts to use the former to study the latter. We started by highlighting Fodor's (2002) working assumption that the two are largely equivalent, and we noted that most investigation into implicit prosodic phrasing has taken place in the context of exploring Fodor's (1998, 2002) IPH. However, the IPH was intended to be an explanation for cross-linguistic patterns in syntactic ambiguity resolution rather than a way to investigate implicit prosody per se, and so work in this area has made it difficult to tease apart questions about the relation between implicit and explicit prosody on the one hand from questions about how readers parse the syntax of printed text on the other (Breen 2014; Jun, 2010). A simple goal of the present study was therefore to explore the relationship between implicit and explicit prosody in a way not so heavily dependent on how speakers/readers resolve syntactic ambiguity, and in a way that takes individual differences into account. A related concern we noted involved the extent to which the prosody obtained from reading aloud—a specific style of explicit prosody—is suitable for the investigation of implicit prosody. In the study presented above, native English speakers participated in two tasks, both involving read-aloud speech, and the output from those two tasks was consistent with a similarity between implicit and explicit prosody. However, this similarity was perhaps not as straightforward as assumed by the IPH. Before concluding, we consider the significance and the limitations of the findings from the two tasks, and the implications for future empirical investigation of implicit prosody.

The goal of the first task (i.e., the production experiment assessing cross-speaker variation in phrase length) was to test for a relationship between individual differences in working memory capacity (WMC) and individual differences in explicit phrase length. Previous work has posited a relationship between WMC and implicit phrase length (e.g., Swets et al., 2007) and to the extent that implicit and explicit prosody are similar, we therefore expected explicit phrase length to also covary with WMC. Consistent with this expectation, we found that speakers with higher WMC were associated with longer prosodic phrases in their explicit productions, accounting for perhaps as much as 20% of the variance across speakers. Furthermore, the production study provided some important data related to the prosodic hierarchy that were not—and probably could not have been—addressed in previous studies of attachment preference like Swets et al., (2007) or Jun and Bishop (2015a). Interestingly, however, there was not a specific level in the prosodic hierarchy that WMC was particularly tied to, as reading spans were most strongly associated with the simple binary distinction between the presence versus absence of a boundary. A corollary of this finding worth highlighting is that subtle, lower-level prosodic boundaries cannot be neglected in analyses of individual differences in phrasing. And, importantly, this fact could only be identified in the context of an explicit phonological model like the one assumed here (Beckman & Pierrehumbert, 1986; Pierrehumbert, 1980), and future research may relate this kind of variation in phrasing to discussions currently taking place in the speech planning literature (e.g., Bishop & Kim, 2018; Kilbourn-Ceron et al., 2020; Krivokapić, 2012; Turk & Shattuck-Hufnagel, 2014; see also Ferreira & Karimi, 2015). For the present purposes, which involve the correlation between explicit and implicit prosodic form, we simply highlight the following: WMC was shown to have the same basic relationship to cross-speaker variation in explicit prosodic phrase length as it has been argued to have with cross-reader variation in implicit prosodic phrase length. This is consistent with a similarity between implicit and explicit prosody, and in this way suggests (but of course, does not on its own prove) that the two may not be fundamentally different in form. And it also needs to be emphasized that the similarity we are referring to here is at the level of broad patterns for individual speakers rather than the kind of highly specific and reliable relationship assumed by the IPH (see Assumption 2a in the introduction, above).

In a second task (and setting aside issues of WMC), we attempted to relate the explicit prosody participants produced in the first task to the implicit prosody they generated during silent reading. In fact, we found participants' explicit phrasing patterns to be systematically related to their implicit phrasing—those who produced longer IPs in the reading-aloud task also tended to identify fewer implicit boundaries while reading silently. It is important to acknowledge that the task employed here has not, to our knowledge, been used before, and it may rely to some extent on judgments that are rather metalinguistic or introspective in nature. However, it is now well-known that untrained participants show some aptitude for identifying prosodic events in auditorily presented speech (Bishop et al., 2020; Hualde et al., 2016) and it is not clear we can dismiss participants' detection of an event simply because the event is an internal one rather than an external one. Many authors have assumed internal speech to be something that language users have some amount of conscious access to (see Clifton, 2015, for a recent review), although accessing it the way participants did in the present study likely required high levels of phonological awareness (e.g., Wagner & Torgesen, 1987).¹² Notably, phonological awareness is known to be dependent on the level of phonological representation (Treiman & Zukowski, 1991) and, as an important predictor of skilled reading in general, is itself subject to individual differences (Quinn et al., 2015; Werfel, 2017). Assuming that labeling phonological structure in this way depends, in part, on phonological awareness, it follows that our implicit boundary identification task was difficult for at least two reasons: (a) because transcribing phonology in general is difficult, and (b) because transcribing *silent* phonology is more difficult yet. It is perhaps therefore not surprising that the relation we found in the second task of this study

between implicit and explicit prosody involved only IPs; IP boundaries mark the highest level in the prosodic hierarchy and are generally the most salient (sometimes marked by actual silent pauses).

This understanding of the task's difficulty may also help to explain why WMC was not a significant predictor of implicit prosodic phrasing in the second task, which, as an anonymous reviewer points out, seems unexpected if implicit and explicit prosody are similar in form. However, recall from the explicit production task that WMC was most strongly associated with the *total number* of prosodic boundaries participants produced—that is, when the distinction between intermediate phrase and IP levels are collapsed—and only weakly associated with boundaries of a particular level. Therefore, if participants were only proficient at identifying IP boundaries in their inner speech, then they were only reporting a subset of the boundaries that are correlated with WMC—which necessarily renders the correlation undetectable. For the present, we therefore suggest that the results of the implicit boundary identification task are consistent with a similarity between implicit and explicit prosody, which, as suggested by Speer and Foltz (2015), may only be detectable in the form of broad tendencies of individual speakers/readers. This conclusion also needs to be tempered with the understanding that (a) the effect size was rather small, and (b) further study is needed to determine the level of sensitivity and reliability this task can achieve. Still, probing listeners internal speech in this way shows some promise as a tool for investigating implicit prosody, and given the well-known difficulty with this object of inquiry, such tools are badly needed.

4 Conclusion

Two basic questions were asked in the present study, motivated by work on implicit prosody, much of which has taken place in the context of the IPH. The first question was whether individual differences in WMC (as estimated by reading spans) predict speakers' explicit prosodic phrase length the way it has been argued to predict implicit prosodic phrase length. The second question (which set aside issues related to WMC) asked whether the individual differences observed in explicit phrase length predicted individual differences in implicit phrasing, determined in a novel implicit boundary identification task. The findings were encouraging in the sense that we were able to provide evidence for a relation between implicit and explicit prosody—and to do so using relatively simple methods and with read-aloud speech. However, they also draw attention to what may be inescapable limitations in studying implicit prosody. As recently discussed by other authors (e.g., Speer & Foltz, 2015), sentence-level prosody, both implicit and explicit, may simply be subject to too much cross-speaker/cross-reader variation to investigate using group-level analyses, or to test predictions framed in terms of highly specific structures. In this way, sentence-level prosody differs considerably from word-level prosody; while an English speaker cannot choose the location of lexical stress on a word, he or she does have options with respect to whether that syllable is marked by a pitch accent—and if so, what tonal shape characterizes that pitch accent. Similarly, lexical and syntactic information do not uniquely determine the location of phrase boundaries, nor what level in the prosodic hierarchy the boundary corresponds to. This problem of cross-speaker variability is further compounded by the fact that within-speaker variability is likely also considerable. What this all means is that we may simply not be able to meaningfully match the prosody of individual sentences produced on one occasion with some measure of implicit prosody generated on some other occasion. Instead, researchers investigating implicit prosody—at least implicit prosody above the word level—may have to limit themselves to questions about “probable” prosodic patterns, and ones that are reader-specific.

In summary, while the wording in the IPH that implicit and explicit prosody are “identical” may be too strong, especially given the amount of variation sentence-level prosody is subject to, it is not clear that the two are fundamentally different in form. Furthermore, in response to some of the

concerns recently discussed in the literature (e.g., Breen, 2014; Jun, 2010), speech from reading-aloud tasks can be used to show this. And while investigating implicit prosody using read-aloud speech does present challenges presented by (a) variability across speakers/readers, (b) variability within-speakers/readers, and (c) inherent differences in the nature of silent versus oral reading, these challenges are not impossible to address.

Acknowledgements

The author is grateful to Nadia Zaki, Eve Elliot, Ting Huang, and Boram Kim for their assistance with the experimental materials and other technical matters. I also thank two anonymous Language and Speech reviewers for their helpful comments and suggestions, as well as audiences at the Framing Speech Workshop at the 8th International Conference on Speech Prosody, the Concordia University (Montreal) Linguistics Department, and the 17th Conference on Laboratory Phonology.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This study was supported in part by Enhanced Grant ENHC-45-52 from the Professional Staff Congress of the City University of New York.

ORCID iD

Jason Bishop  <https://orcid.org/0000-0001-7346-9836>

Notes

1. For a thorough and fairly recent review, see Breen (2014)
2. Though not crucial to the present purposes, the authors actually found that domain-general working memory resources were implicated in the effects they reported. However, verbal working memory appeared to be more important than spatial WMC in their study, and it has also been found to be a significant predictor of attachment preference on its own in other work (e.g., Jun and Bishop, 2015a, Expt. 1).
3. In general, the examples of phrasing patterns here make reference to languages that have the linear surface word order NP1 NP2 RC (e.g., “. . . *the servant of the actress who was on the balcony*”), as does English, German, and Spanish. But the prediction that the IPH makes about phrasing does not depend on this word order. For example, in Korean and Japanese, the word order (relative to the English order) is RC NP2 NP1, and so the predicted high attachment phrasing is (RC)/(NP2 NP1) and the low attachment phrasing (RC NP2)/(NP1). In both word orders what is prosodically crucial is whether the RC is phrased together with NP2 or separately from both NPs.
4. For example, it is now understood that prominence (i.e., accentuation) needs to be taken into account in at least some languages (Carlson & Tyler, 2017; Gryllia & Kügler, 2010; Jun & Bishop, 2015a; Lee & Watson, 2011; Schafer et al., 1996). It also seems clear that implicit prosody’s influence interacts with parsing biases that are purely syntactic in nature (Kentner, 2017; Watson & Gibson, 2004), the details of which may depend on stage of processing (Hwang & Schafer, 2009; Traxler, 2007/2009;) and perhaps even developmental differences across the lifespan (Payne et al., 2014). Finally, Grillo (2012) and Grillo and Costa (2014) argue that cross-linguistic differences in attachment preference correlate with grammatical factors not previously considered (and which do not necessarily require reference to prosody).
5. While there does not seem to be a study investigating this particular question from a phonological perspective, it is worth pointing out that work in the domain of speech production planning provides some basis for predictions regarding a relationship between WMC and spoken phrase length, although it requires tying together two separate strands of research. The first strand is work demonstrating that speakers’ planning unfolds in (minimally) intonational-phrase-sized chunks, and thus that higher-level prosodic units are potential candidates for planning units (Ferreira, 1993; Keating & Shattuck-Hufnagel, 2002; see Shattuck-Hufnagel, 2015, for a recent review). The second strand is work that seems to indicate

that the size of speakers' preferred planning units—and thus preferred prosodic units—may vary due to both external pressures related to speaking conditions, as well as internal pressures related to processing resources (Ferreira & Swets, 2002; Fuchs et al., 2013; Krivokapić 2007a, 2007b, 2012; Petrone et al., 2011; Swets et al., 2013; Swets et al., 2014; Wagner et al., 2010; see also recent discussion in Ferreira & Karimi, 2015). While these two strands of research factored prominently into the present author's thinking (as well as an earlier version of this manuscript), for both concision and clarity, the present discussion focuses on the issues discussed in the implicit prosody literature.

6. See archive, still accessible as of 20 August 2020, at <https://obamawhitehouse.archives.gov/briefing-room/weeklyaddress>.
7. Indeed they were somewhat higher than some of these studies report. This is likely to be in part the result of the style of speech elicited (fairly carefully read speech, recorded in a laboratory) as well as the methodology (transcribers were making decisions about a limited set of pre-determined potential boundaries, identified by the first-pass transcribers).
8. As mentioned above, no previous research seems to have explored the issue in an individual differences context like the present study, but see Krivokapić (2012), who explicitly predicts (based on phonetic measures of speech planning) that reduced processing resources (such as low WMC) should affect both lower and higher levels of structure.
9. Given the miniscule number of filled pauses produced (a total of four, produced by two of the 100 speakers), they are omitted from the discussion.
10. A reviewer asks whether the passages utilized in the present study were of sufficient length to allow for reasonable estimates of phrasing behavior. Methodological considerations in the present study included the need to limit the duration of the experiment for participants (who completed multiple tasks), and for expert annotators (who would be employed to transcribe the phonological structure of the samples). Split-half reliability coefficients (with Spearman-Brown corrections) indicated that the explicitly read passage allowed for reasonably stable estimates of the overall number of breaks ($r = .79$) and the number of IPs they produced ($r = .80$). However, the stability of IP-internal ip boundaries less so ($r = .42$), and so the conclusions we draw below will be strongest for boundaries in general and boundaries at the higher, IP level. For the implicit task, to be described further below, split-half reliability for boundaries identified by participants in the silently read passage was shown to be relatively high ($r = .88$).
11. If the two outlying data points (representing the lowest end of RSPAN scores) are dropped, the R^2 values in Figure 3 do in fact decrease to $R^2 = .039$ (for ips), $R^2 = .069$ (for IPs), and $R^2 = .153$ (for total number of boundaries). However, none of the crucial statistical patterns presented below depend on the presence or absence of these two data points.
12. The author thanks Janet Fodor for discussion of this issue. Fodor (2002) noted early on the possibility that individual readers may differ in the “strength” of their implicit prosodic representations, which is subtly different from the issue of phonological awareness just alluded to. This possibility does not seem to have been pursued in subsequent research, but see Jun and Bishop (2015a), who appeal to differences in the encoding of, or memory for, implicit accentuation patterns to account for interindividual differences in priming patterns.

References

- Alda, A. (2006). *Never have your dog stuffed: And other things I've learned*. Random House.
- Bader, M. (1998). Prosodic influences on reading syntactically ambiguous sentences. In J. Fodor & F. Ferreira (Eds.): *Reanalysis in sentence processing* (pp. 1–46). Kluwer. doi:10.1007/978-94-015-9070-9_1
- Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The autism-spectrum quotient (AQ): Evidence from Asperger syndrome/high-functioning autism, males, females, scientists and mathematicians. *Journal of Autism & Developmental Disorders*, 31(1), 5–17. doi:10.1023/A:1005653411471
- Beckmann, M. & Ayers Elam, G. (1997). Guidelines for ToBI labeling (Version 3). Unpublished ms. Ohio State University.
- Beckman, M. & Hirschberg, J. (1994). The ToBI annotation conventions. Unpublished ms. Ohio State University, Columbus, OH.

- Beckman, M. & Pierrehumbert, J. (1986). Intonational structure in Japanese and English. *Phonology Yearbook* 3, 255–309.
- Bergmann, A., Armstrong, M., & Maday, K. (2008). Relative clause attachment in English and Spanish: A production study. *Proceedings of Speech Prosody 2008*, 507–508.
- Bishop, J., Kuo, G., & Kim, B. (2020). Phonology, phonetics, and signal-extrinsic factors in the perception of prosodic prominence: Evidence from Rapid Prosody Transcription. *Journal of Phonetics*, 82, 100977. doi.org:10.1016/j.wocn.2020.100977
- Bookheimer, S., Zeffiro, T., Blaxton, T., Gaillard, W., & Theodore, W. (1995). Regional cerebral blood flow during object naming and word reading. *Human Brain Mapping*, 3(2), 93–106. doi:10.1002/hbm.460030206
- Bortfeld, H., Leon, S., Bloom, J., Schober, M., & Brennan, S. (2001). Disfluency rates in conversation: effects of age, relationship, topic, role, and gender. *Language and Speech*, 44(2), 123–147. doi:10.1177/00238309010440020101
- Breen, M. (2014). Empirical investigations of the role of implicit prosody in sentence processing. *Language and Linguistics Compass*, 8(2), 37–50. doi:10.1111/lnc3.12061
- Breen, M., Dilley, L., Kraemer, J., & Gibson, E. (2012). Inter-transcriber reliability for two systems of prosodic annotation: ToBI (Tones and Break Indices) and RaP (Rhythm and Pitch). *Corpus Linguistics and Linguistic Theory*, 8(2), 277–312. doi:10.1515/cllt-2012-0011
- Breen, M., Fedorenko, E., Wagner, M., & Gibson, E. (2010). Acoustic correlates of information structure. *Language and Cognitive Processes*, 25(7), 1044–1098. doi:10.1080/01690965.2010.504378
- Breen, M., Watson, D., & Gibson, E. (2011). Intonational phrasing is constrained by meaning, not balance. *Language and Cognitive Processes*, 26(10), 1532–1562. doi:10.1080/01690965.2010.508878
- Büring, D. (2016) *Intonation and meaning*. Oxford University Press.
- Calhoun, S. (2012). The theme/rheme distinction: Accent type or relative prominence? *Journal of Phonetics*, 40, 329–349. doi:10.1016/j.wocn.2011.12.001
- Carlson, K., & Tyler, J. (2017). Accents, not just prosodic boundaries, influence syntactic attachment. *Language and Speech*. Published online ahead of print. doi:10.1177/0023830917712282
- Carreiras, M., & Clifton, C. (1993). Relative clause interpretation preferences in Spanish and English. *Language and Speech*, 36, 353–372. doi:10.1177/002383099303600401
- Clifton, C. (2015). The roles of phonology in silent reading: A selective review. In L. Frazier & E. Gibson (Eds.): *Explicit and implicit prosody in sentence processing: Studies in honor of Janet Dean Fodor (Studies in Theoretical Psycholinguistics 46)* (pp 161–176). Springer. doi:10.1007/978-3-319-12961-7
- Cole, J., Hualde, J. I., Eager, C., & Mahrt, T. (2015). On the prominence of accent in stress reversal. In: The Scottish Consortium for ICPHS 2015 (Ed.), *Proceedings of the 18th International Congress of Phonetic Sciences* (pp. 1–5). Glasgow, UK: the University of Glasgow, <http://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS0771.pdf>
- Cole, J., Mahrt, T., & Roy, J. (2017). Crowd-sourcing prosodic annotation. *Computer Speech & Language*, 45, 300–325. doi:10.1016/j.csl.2017.02.008
- Cole, J., Mo, Y., & Baek, S. (2010). The role of syntactic structure in guiding prosody perception with ordinary listeners and everyday speech. *Language and Cognitive Processes*, 25(7), 1141–1177. doi:10.1080/01690960903525507
- Cole, J., Mo, Y., & Hasegawa-Johnson, M. (2010). Signal-based and expectation-based factors in the perception of prosodic prominence. *Laboratory Phonology 1*(1), 425–452. doi:10.1515/labphon.2010.022
- Conway, A., Kane, M., Bunting, M., Hambrick, D., Wilhelm, O., & Engle, R. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review* 12, 769–786. doi.org/10.3758/BF03196772
- Cooper, W., & Paccia-Cooper, J. (1980). *Syntax and speech*. Harvard University Press.
- 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. doi:10.1016/0010-0277(88)90004-2
- Daneman, M. (1991). Working memory as a predictor of verbal fluency. *Journal of Psycholinguistic Research*, 20(6), 445–464. doi:10.1007/BF01067637

- Daneman, M., & Carpenter, P. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning & Verbal Behavior*, *19*, 450–466. doi:10.1016/S0022-5371(80)90312-6
- Ehrlich, K., Fernández, E., Fodor, J. D., Stenshoel, E., & Vinereanu, M. (1999, March 18–20). *Low attachment of relative clauses: New data from Swedish, Norwegian, and Romanian. Poster presented at the 12th annual CUNY Conference on Human Sentence Processing, New York, NY.*
- Engelhardt, P., Nigg, J., & Ferreira, F. (2013). Is the fluency of language outputs related to individual differences in intelligence and executive function? *Acta Psychologica*, *144*, 424–432. doi:10.1016/j.actpsy.2013.08.002
- Fernández, E. (2003). *Bilingual sentence processing: Relative clause attachment in English and Spanish*. Amsterdam: John Benjamins. doi:10.1075/lald.29
- Fernández, E., & Bradley, D. (1999, March 18–20). *Length effects in the attachment of relative clauses in English. Poster presented at the 12th annual CUNY Conference on Human Sentence Processing, New York.*
- Ferreira, F. (1993). Creation of prosody during sentence production. *Psycholinguistic Review*, *100*, 233–253. doi:10.1037/0033-295X.100.2.233
- 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. doi:10.1006/jmla.2001.2797
- Ferreira, F., & Karimi, H. (2015). Prosody, performance, and cognitive skill: Evidence from individual differences. In L. Frazier & E. Gibson (Eds.): *Explicit and implicit prosody in sentence processing: Studies in honor of Janet Dean Fodor (Studies in Theoretical Psycholinguistics 46)* (pp. 119–132). Springer. doi:10.1007/978-3-319-12961-7
- Fodor, J. D. (1998). Learning to parse? *Journal of Psycholinguistic Research*, *27*, 285–319. doi:10.1023/A:1023258301588
- Fodor, J. D. (2002). Prosodic disambiguation in silent reading. In M. Hirotani (Ed.), *Proceedings of the North East Linguistics Association* (pp. 112–132). (NELS) 32.
- Fougeron, C., & Jun, S.-A. (1998). Rate effects on French intonation: Phonetic realization and prosodic organization. *Journal of Phonetics*, *26*(1), 45–70. doi:10.1006/jpho.1997.0062
- Fox Tree, J. (1995). The effects of false starts and repetitions on the processing of subsequent words in spontaneous speech. *Journal of Memory and Language*, *34*, 709–738. doi:10.1006/jmla.1995.1032
- Frazier, L. (1979). *On comprehending sentences: Syntactic parsing strategies*. [Unpublished doctoral dissertation]. University of Connecticut.
- Fuchs, S., Petrone, C., Krivokapić, K., & Hoole, P. (2013). Acoustic and respiratory evidence for utterance planning in German. *Journal of Phonetics*, *41*, 29–47. doi:10.1016/j.wocn.2012.08.007
- Gee, P., & Grosjean, F. (1983). Performance structures: A psycholinguistic and linguistic appraisal. *Cognitive Psychology*, *15*, 411–458. doi:10.1016/0010-0285(83)90014-2
- Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition*, *68*, 1–76. doi:10.1016/S0010-0277(98)00034-1
- Golden, C. (1978). *Stroop color and word test: A manual for clinical and experimental issues*. Stoelting.
- Grillo, N. (2012). Local and universal. In V. Bianchi & C. Chesì (Eds.): *Enjoy Linguistics! Papers offered to Luigi Rizzi on the occasion of his 60th birthday* (pp. 234–245). CISCL Press.
- Grillo, N., & Costa, J. (2014). A novel argument for the universality of parsing principles. *Cognition*, *133*(1), 156–187. doi:10.1016/j.cognition.2014.05.019
- Griffin, Z. (2001). Gaze durations during speech reflect word selection and phonological encoding. *Cognition*, *82*(1), B1–B14. doi:10.1016/S0010-0277(01)00138-X
- Gryllia, S., & Kügler, F. (2010) What does prosody tell us about relative clause attachments in German? *Proceedings of Speech Prosody 2010*, 100927: 1–4.
- Hemforth, B., Fernandez, S., Clifton, C., Frazier, L., Konieczny, L., & Walter, M. (2015). Relative clause attachment in German, English, Spanish and French: Effects of position and length. *Lingua*, *166*(A), 43–64. doi:10.1016/j.lingua.2015.08.010
- Hirotani, M., Frazier, L., & Rayner, K. (2006). Punctuation and intonation effects on clause and sentence wrap-up: Evidence from eye movements. *Journal of Memory and Language*, *54*, 425–43. doi:10.1016/j.jml.2005.12.001

- Hualde, J.I., Cole, J., Smith, C., Eager, C., Mahrt, T., Napoleão, de, & Souza, R. (2016). The perception of phrasal prominence in English, Spanish and French conversational speech. *Proceedings of Speech Prosody 8*, 459–463.
- Hwang, H., & Schafer, A. (2009). Constituent length affects prosody and processing for a dative NP ambiguity in Korean. *Journal of Psycholinguistic Research*, 38, 151–175. doi:10.1007/s10936-008-9091-1
- Ito, K., & Speer, S. (2006). Using interactive tasks to elicit natural language dialogue. In P. Augurzyk & D. Lenertova (Eds.), *Methods in Empirical Prosody Research*. (pp. 229–258). Mouton de Gruyter.
- Jun, S.-A. (1993). The phonetics and phonology of Korean prosody. [unpublished doctoral dissertation]. The Ohio State University, Columbus, Ohio.
- Jun, S.-A. (2003a). Prosodic phrasing and attachment preferences. *Journal of Psycholinguistic Research*, 32(2), 219–249. doi:10.1023/A:1022452408944
- Jun, S.-A. (2003b). The effect of phrase length and speech rate on prosodic phrasing. In: M.J. Solé, D. Recasens, & J. Romero (Eds.), *Proceedings of the 15th International Congress of Phonetic Sciences* (pp. 483–486). www.internationalphoneticassociation.org/icphs/icphs2003
- Jun, S.-A. (2007). The intermediate phrase in Korean intonation: Evidence from sentence processing. In C. Gussenhoven & T. Riad (Eds.), *Tones and tunes: Studies in word and sentence prosody* (pp. 143–167). Mouton de Gruyter.
- Jun, S.-A. (2010). The implicit prosody hypothesis and overt prosody in English. *Language and Cognitive Processes*, 25(7), 1201–1233. doi:10.1080/01690965.2010.503658
- Jun, S.-A., & Bishop, J. (2015a). Prominence in relative clause attachment: Evidence from prosodic priming. In L. Frazier & E. Gibson (Eds.), *Explicit and implicit prosody in sentence processing: Studies in honor of Janet Dean Fodor (Studies in Theoretical Psycholinguistics 46)* (pp. 217–240). Springer. doi:10.1007/978-3-319-12961-7
- Jun, S.-A., & Bishop, J. (2015b). Priming implicit prosody: Prosodic boundaries and individual differences. *Language and Speech*, 58(4), 459–473. doi:10.1177/0023830914563368
- Jun, S.-A., & Kim, S. (2004). Default phrasing and attachment preferences in Korean. *Proceedings of Interspeech – International Conference on Spoken Language Processing, Jeju, Korea*.
- Jun, S.-A., & Koike, C. (2003). Prosody and attachment preference in Japanese: Production and perception. *Proceedings of the International Workshop on Prosodic Interfaces, Nantes, France*.
- Jun, S.-A., & Shilman, M. (2008). Default phrasing and English relative clause attachment data. In P. Barbosa, S. Madureira, & C. Reis (Eds.), *Proceedings of Speech Prosody 2008* (pp. 469–472). www.isca-speech.org/archive/sp2008
- Kamide, Y., & Mitchell, D. (1997). Relative clause attachment: Non-determinism in Japanese parsing. *Journal of Psycholinguistic Research*, 26, 247–254. doi:10.1023/A:1025017817290
- Keating, P., & Shattuck-Hufnagel, S. (2002). A prosodic view of word form encoding for speech production. *UCLA Working Papers in Phonetics*, 101, 112–156.
- Kentner, G. (2017). Rhythmic parsing. *The Linguistic Review*, 34(1), 123–155. https://doi.org/10.1515/trl-2016-0005
- Kilbourn-Ceron, O., Clayards, M., & Wagner, M. (2020). Predictability modulates pronunciation variants through speech planning effects: A case study on coronal stop realizations. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, 11(1), 1–28.
- Kimball, J. (1973). Seven principles of surface structure parsing in natural language. *Cognition*, 2, 15–47. doi:10.1016/0010-0277(72)90028-5
- Krivokapić, J. (2007a). *The planning, production, and perception of prosodic structure*. [unpublished doctoral dissertation]. University of Southern California.
- Krivokapić, J. (2007b). Prosodic planning: Effects of phrasal length and complexity on pause duration. *Journal of Phonetics*, 35, 162–179. doi:10.1016/j.wocn.2006.04.001
- Krivokapić, J. (2012). Prosodic planning in speech production. In S. Fuchs, M. Wehrich, D. Pape, & P. Perrier (Eds.), *Speech planning and dynamics* (pp. 157–190). Peter Lang. doi:10.3726/978-3-653-01438-9
- Ladd, R. (2008). *Intonational phonology* (2nd Edition). Cambridge University Press.

- Lee, E.-K., & Watson, D. (2011). Effects of pitch accents in attachment ambiguity resolution. *Language and Cognitive Processes, 26*, 262–297. doi: 10.1080/02F01690965.2010.491650
- Lickley, R. (2015). Fluency and disfluency. In M. Redford (Ed.), *The Blackwell Handbook of Speech Production* (pp. 445–474). Wiley-Blackwell.
- Matuschek, H., Kliegl, R., Vasishth, S., Baayen, R. H., & Bates, D. (2017). Balancing type I error and power in linear mixed models. *Journal of Memory and Language, 94*, 305–315. doi:10.1016/j.jml.2017.01.001
- Maynell, L. (1999, March 18–20). *Effect of pitch accent placement on resolving relative clause ambiguity in English*. Poster presented at the 12th CUNY Conference on Human Sentence Processing, New York, NY.
- Nespor, I., & Vogel, I. (1986). *Prosodic Phonology*. Foris Publications.
- Oswald, F., McAbee, S., Redick, T., & Hambrick, D. (2015). The development of a short domain-general measure of working memory capacity. *Behavior Research Methods, 47*(4), 1343–1355. doi:10.3758/s13428-014-0543-2
- Payne, B., Grison, S., Gao, X., Christianson, K., Morrow, D., & Stine-Morrow, E. (2014). Aging and individual differences in binding during sentence understanding: Evidence from temporary and global syntactic attachment ambiguities. *Cognition, 130*(2), 157–173. doi:10.1016/j.cognition.2013.10.005
- Petrone, C., Fuchs, S., & Krivokapić, J. (2011). Consequences of working memory differences and phrasal length on pause duration and fundamental frequency. In Y. Laprie & I. Steinder (Eds.), *Proceedings of the 9th International Seminar on Speech Production* (pp. 393–400).
- Pierrehumbert, J. (1980). *The phonology and phonetics of English intonation*. [unpublished doctoral dissertation]. Massachusetts Institute of Technology.
- Pitrelli, J., Beckman, M., & Hirschberg, J. (1994). Evaluation of prosodic transcription labeling reliability in the ToBI framework. *Proceedings of the 3rd International Conference on Spoken Language Processing*. (pp. 123–126).
- Pynte, J., & Prieur, B. (1996). Prosodic breaks and attachment decisions in sentence processing. *Language and Cognitive Processes, 11*, 165–191. doi:10.1080/016909696387259
- Quinn, J., Spencer, M., & Wagner, R. (2015). Individual differences in phonological awareness and their role in learning to read. In P. Afflerbach (Ed.), *Handbook of individual differences in reading: Reader, text, and context* (pp. 80–92). Routledge.
- Rasinski, T. (2006). Reading fluency instruction: Moving beyond accuracy, automaticity, and prosody. *The Reading Teacher, 59*(7), 704–706. doi:10.1598/RT.59.7.10
- Rayner, K., Pollatsek, A., Ashby, J., & Clifton, C. (2012). *The psychology of reading*. Psychology Press.
- Roy, J., Cole, J., & Mahrt, T. (2017). Individual differences and patterns of convergence in prosody perception. *Laboratory Phonology, 8*(1), 22. doi:10.5334/labphon.108
- Schafer, A., Carter, J., Clifton, C., & Frazier, L. (1996). Focus in relative clause construal. *Language and Cognitive Processes, 11*, 135–163. doi:10.1080/016909696387240
- Schremm, A., Horne, M., & Roll, M. (2015). Brain responses to syntax constrained by time-driven implicit prosodic phrases. *Journal of Neurolinguistics, 35*, 68–84. doi:10.1016/j.jneuroling.2015.03.002
- Selkirk, E. (2000). The interaction of constraints on prosodic phrasing. In M. Horne (Ed.), *Prosody: theory and experiment (Studies presented to Gösta Bruce)* (pp. 231–262). Kluwer Academic. doi:10.1007/978-94-015-9413-4
- Shattuck-Hufnagel, S. (2015). Prosodic frames in speech production. In M. Redford (Ed.), *The handbook of speech production* (pp. 419–444). Wiley-Blackwell.
- Shriberg, E. (1994). *Preliminaries to a theory of speech disfluencies*. [unpublished doctoral dissertation], University of California, Berkeley.
- Silverman, K., Beckman, M., Pitrelli, J., Ostendorf, M., Wightman, C., Price, P., Pierrehumbert, J., & Hirschberg, J. (1992). ToBI: A standard for labeling English prosody. *Proceedings of International Conference on Spoken Language Processing, 2*, (pp. 867–870).
- Slowiaczek, M., & Clifton, C. (1980). Subvocalization and reading for meaning. *Journal of Verbal Learning & Verbal Behavior, 19*(5), 573–582. doi:10.1016/S0022-5371(80)90628-3
- Snedeker, J., & Trueswell, J. (2003). Using prosody to avoid ambiguity: Effects of speaker awareness and referential context. *Journal of Memory and Language, 48*, 103–130. doi:10.1016/S0749-596X(02)00519-3

- Speer, S., Kjelgaard, M., & Debroth, K. (1996). The influence of prosodic structure on the resolution of temporary syntactic closure ambiguities. *Journal of Psycholinguistic Research*, 25(2), 249–271. doi:10.1007/BF01708573
- Speer, S., & Foltz, A. (2015). The implicit prosody of corrective contrast primes appropriately intonated primes (for some readers). In L. Frazier & E. Gibson (Eds.), *Explicit and implicit prosody in sentence processing: Studies in honor of Janet Dean Fodor (Studies in Theoretical Psycholinguistics 46)* (pp. 263–285). Springer. doi:10.1007/978-3-319-12961-7
- Steinhauer, K. 2003. Electrophysiological correlates of prosody and punctuation. *Brain and Language*, 86, 142–164. doi:10.1016/S0093-934X(02)00542-4
- Steinhauer, K., & Freiderici, A. (2001). Prosodic boundaries, comma rules, and brain responses: The closure positive shift in ERPs a universal marker for prosodic phrasing in listeners and readers. *Journal of Psycholinguistic Research*, 30, 267–295. doi:10.1023/A:1010443001646
- Stroop, J. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 18, 643–662. doi:10.1037/h0054651
- Swets, B., Desmet, T., Hambrick, D., & Ferreira, F. (2007). The role of working memory in syntactic ambiguity resolution: A psychometric approach. *Journal of Experimental Psychology: General*, 136, 64–81. doi:10.1037/0096-3445.136.1.64
- Swets, B., Jacovina, M., & Gerrig, R. (2013). Effects of conversational pressures on speech planning. *Discourse Processes*, 50, 23–51. doi:10.1080/0163853X.2012.727719
- Swets, B., Jacovina, M., & Gerrig, R. (2014). Individual differences in the scope of speech planning: Evidence from eye movements. *Language and Cognition*, 6, 12–44. doi:10.1017/langcog.2013.5
- Syrdal, A., & McGory, J. (2000). Inter-transcriber reliability of ToBI prosodic labeling. *Proceedings of the 6th International Conference on Spoken Language Processing*, (pp. 235–238).
- Tanner, J., Sonderegger, M., & Wagner, M. (2017). Production planning and coronal stop deletion in spontaneous speech. *Laboratory Phonology*, 8(1), 15. doi:10.5334/labphon.96
- Traxler, M. (2007). Working memory contributions to relative clause attachment processing: A hierarchical linear modeling analysis. *Memory and Cognition*, 35, 1107–1121. doi:10.3758/BF03193482
- Traxler, M. (2009). A hierarchical linear modeling analysis of working memory and implicit prosody in the resolution of adjunct attachment ambiguity. *Journal of Psycholinguistic Research*, 38, 491–509. doi:10.1007/s10936-009-9102-x
- Treiman, R., & Zukowski, A. (1991). Levels of phonological awareness. In S.A. Brady & D.P. Shankweiler (Eds.), *Phonological processes in literacy: A tribute to Isabelle Y. Liberman* (pp. 67–83). Lawrence Erlbaum Associates.
- Turk, A., & Shattuck-Hufnagel, S. (2014). Timing in talking: What is it used for, and how is it controlled? *Philosophical Transactions of the Royal Society B*, 369, 20130395. doi:10.1098/rstb.2013.0395
- Turnbull, R., Royer, A., Ito, K., & Speer, S. (2017). Prominence perception is dependent on phonology, semantics, and awareness of discourse. *Language, Cognition and Neuroscience*. [Published online ahead of print]. doi:10.1080/23273798.2017.1279341
- Wagner, V., Jescheniak, J., & Schriefers, H. (2010). On the flexibility of grammatical advance planning during sentence production: Effects of cognitive load on multiple lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36, 423–440. doi:10.1037/a0018619
- Wagner, R. K., & Torgesen, J. K. (1987). The nature of phonological processing and its causal role in the acquisition of reading skills. *Psychological Bulletin*, 101, 192–212.
- Watson, D., & Gibson, E. (2004). The relationship between intonational phrasing and syntactic structure in language production. *Language and Cognitive Processes*, 19, 713–755. doi:10.1080/01690960444000070
- Werfel, K. (2017). Phonetic transcription training improves adults' explicit phonemic awareness: Evidence from undergraduate students. *Communication Disorders Quarterly* 39(1), 281–287. <https://doi.org/10.1177/1525740117702456>
- Yoon, T., Chavarria, S., Cole, J., & Hasegawa-Johnson, M. (2004). Inter-transcriber reliability of prosodic labeling on telephone conversation using ToBI. *8th International Conference on Spoken Language Processing*, (pp.2729–2732).

Appendix

Complete passage read by participants in the reading-aloud task (explicit prosody)

Our patch of tangled yard was an exotic foreign country. I had spent so much of my life in dark theaters and dim hotel rooms, where the only thing green was the peeling paint on the walls, that this seemed perfectly natural to me. This was where I had my first bite of mud pie; where I set up a card table and mixed household chemicals, toothpaste, and my mother's face powder, doing what I called "experiments". Now that I was sick I wasn't allowed out of the house, but when I stood on the headboard of my bed I could look through a high window into the backyard and see the concrete wall I used to climb over with my friends. We'd sneak under the cover of the banana trees and light matches I had stolen from my mother's kitchen. I could just reach the big box of wooden matchsticks she kept on the top shelf of the old four-legged gas stove.

Complete passage read by participants in the silent-reading task (implicit prosody)

I had come down with polio a couple of weeks after starting second grade. My parents had enrolled me in a military academy, possibly acting once again on their weird fetish for uniforms. For the first two weeks of school, I was introduced to the fine points of drilling, marching, and standing guard. Standing guard involved walking in circles for an hour or two and being able to recite, if asked, a list of ten special orders. There were rumors among my classmates about what would happen to you if you got any of the orders wrong or in the wrong sequence. Older boys would come up to you and fire off questions as if they were the officer of the day. What's order number three? What's order number five? This was an interesting departure from standing in the wings watching Miss Anderson change into her costume.

It has surprised me as I've traveled from town to town how often people have asked me with real interest about what happened to the stuffed dog. I have no idea what happened to it. I was so disappointed in how he looked when he came back from the taxidermist that I had no sentimental attachment to the thing at all. As time went on, I became far more interested in what the stuffed dog had come to stand for: hanging on to the past, standing in place, and not moving on. Getting him out of my sight and out of my head would have been fine with me.