Prosody indexes both competence and performance

Duane G. Watson¹ | Cassandra L. Jacobs² | Andrés Buxó-Lugo³

Abstract

Prosody is an important feature of language that conveys a wide range of information. However, prosody is widely considered to be a difficult domain of study within the language sciences. One consequence of this is that existing grammatical theories of prosody fail to explain prosodic choices that seem to arise from non-linguistic cognitive demands, such as communicative context, top-down expectations, and recent articulatory and acoustic experience. We provide an account of some of these phenomena and argue that linguistic theories that do not incorporate these factors into models of prosody are likely to mischaracterize its role in language.

This article is categorized under:
- Linguistics > Language in Mind and Brain
- Psychology > Language
- Linguistics > Linguistic Theory

KEYWORDS
- competence, language, performance, prosody

1 INTRODUCTION

A challenge for the language sciences is characterizing the representations that underlie prosody as well as building theories of how prosody functions in communication. Prosody can be defined as aspects of the acoustic signal that convey information independently of the segmental information in a sentence. It is typically realized acoustically by changes in pitch (intonation), intensity (volume), and duration (timing), but subjectively, prosody is experienced as emphasis on a word, pauses in an utterance, and changes in intonation, to name just a few examples. It is impossible to speak without prosody: it exists in every language, signed or oral, and is even generated during reading (e.g., Breen, 2014; Fodor, 2002). Prosodic form (e.g., pitch) regularly correlates with other aspects of linguistic structure that are critical for determining what a speaker means. These facets include syntax, semantics, segmental phonology, and pragmatics (Büring, 2007; Steedman, 2000; Wagner & Watson, 2010, inter alia). Prosody also conveys paralinguistic information, such as happiness, fear, or frustration (Crystal & Quirk, 1964). Despite prosody’s important role in language, there is disagreement about the acoustic correlates of prosodic phenomena (e.g., accenting; Beckman, 1986; Breen, Fedorenko, Wagner, & Gibson, 2010; Eady, Cooper, Klouda, Mueller, & Lotts, 1986; Kochanski, Grabe, Coleman, & Rosner, 2005; Ladd & Morton, 1997; Sluijter & van Heuven, 1996) or the meanings that prosody can be used to convey (compare, for example, Pierrehumbert and Hirschberg (1990) to Aylett and Turk (2004) to Selkirk (1984) to Wagner (2016)).

Typically, prosody is studied through the lens of linguistic competence, or idealized representations that link form to meaning. Traditionally, the distinction between competence and performance (Chomsky, 1965) has been used to motivate the study of linguistic knowledge independent of language use. However, in this paper, we argue that it is critical to study prosodic knowledge at the same time as prosodic use. We argue that a necessary component of understanding prosodic structure is...
understanding the psychological processes that generate and interpret prosody. The structure of the paper continues as follows: First, we discuss the types of linguistic information that prosody conveys. We then review work showing that the perception and production of prosody appears to be sensitive to cognitive constraints and communicative factors, as well as traditional linguistic factors, such as syntactic structure and information status. We then outline why understanding prosodic structure requires models that tightly integrate grammatical representations with processing and contextual constraints.

2 | WHAT INFORMATION IS THERE IN THE PROSODIC SIGNAL?

There is a great deal of evidence suggesting that prosody conveys information about other levels of linguistic representation. In English, prosody conveys syntactic, discourse, and pragmatic information (Shattuck-Hufnagel & Turk, 1996; Wagner & Watson, 2010). For example, we know that speakers tend to place prosodic breaks at major syntactic boundaries. Consider the sentence below:

1. Andrés decided to try a special dish from Rochester called a garbage plate // but he regretted ordering it as soon as he saw it.

Most speakers are likely to place a break (or pause) between “plate” and “but” because this is a boundary between two syntactic clauses (Breen, Watson, & Gibson, 2011; Cooper & Paccia-Cooper, 1980; Gee & Grosjean, 1983; Watson & Gibson, 2004). Listeners take advantage of this link between syntax and prosody in production, by using boundary placement to infer the speaker's intended syntactic structure (Snedeker & Trueswell, 2003; Speer, Warren, & Schafer, 2011). Consider for example the following Groucho Marx joke from the movie Animal Crackers:

2. One morning I shot an elephant in my pajamas. How he got in my pajamas, I don't know.

The joke's punchline depends on a syntactic ambiguity, where the listener incorrectly assumes that Groucho, and not the elephant is wearing pajamas. This ambiguity can be resolved with an intonational boundary. A boundary after “shot” creates a bias towards an interpretation in which the elephant is wearing pajamas. A boundary after “elephant” creates an interpretation in which Groucho is wearing pajamas. Critically, prosody is useful to a listener precisely because it conveys information about syntactic structure.

We see similar types of mappings between prosodic structure and other levels of linguistic representation. In English, new and unpredictable words tend to be marked with pitch accents, which are acoustic prominences marked by foregrounding words with fundamental frequency (F0), duration, and intensity (Chafe, 1987; Fowler & Housum, 1987; Prince, 1981). Similarly, pragmatic information, such as marking a sentence as a question vs. a statement, can be conveyed prosodically (Gunlogson, 2003). The difference between a statement like “You're leaving me.” and a question like “You're leaving me?” is signaled by prosody, and not the individual words in the sentence.

This mapping between prosody and other aspects of linguistic structure seems to be straightforward: a given prosodic cue maps onto a specific pragmatic meaning, syntactic structure, or discourse status. However, as we discuss below, these examples belie the complexity of this mapping.

3 | PROSODY AND COMPETENCE

Linguists have described this relationship between prosody and other levels of linguistic representation by using formal approaches. Some theories have proposed linguistic rules or constraints that describe the relationship between intonation and meaning (Gussenhoven, 1984; Pierrehumbert & Hirschberg, 1990; Rochemont, 1986; Schmerling, 1976; Selkirk, 1984); prosody and syntax (Nespor & Vogel, 1986; Selkirk, 1995; Truckenbrodt, 1999); prosody and morphology (McCarthy & Prince, 1993); and prosody and segmental phonology (Beaver, 1968; Chomsky & Halle, 1968; Ohala & Kawasaki, 1984), among others. These approaches have provided important insights into prosodic choices and acoustic correlates of prosody. One common idea among these approaches is that the mapping between prosody and other aspects of grammar can be formalized as a system of rules and constraints to explain how prosody is realized depending on other linguistic factors. On one hand, this approach has been quite useful in uncovering some of the factors that seem to effect prosody. On the other hand, these approaches fall short of explaining some speaker and listener prosodic preferences, which we outline below.
4 | PROSODY AND PERFORMANCE

Some findings from the literature suggest that the perception of prosody is driven by several cognitive and contextual factors that are at least partly independent of linguistic ones. It is often acknowledged that performance factors will affect the production and comprehension of prosody (Ferreira & Karimi, 2015; Turk & Shattuck-Hufnagel, 2014). In fact, in some cases, studies have revealed whether a certain phenomenon can be best explained by competence or performance (Wagner & Klassen, 2015). However, competence and performance are not always separable. Below, we discuss experiments showing that (a) the perception of syntactic boundaries is influenced by expectations; (b) the communicative goals of the speaker strongly influence the production of acoustic prominence; and (c) production-internal factors such as difficulties in phonological sequencing influence speech rate and whole-word duration.

4.1 | Prosody and expectations

The traditional approach to studying prosodic comprehension has consisted of building prosodic inventories that define different prosodic categories based on their acoustic manifestations, and mapping this set of cues to a meaning. For example, an acoustic disjunction in speech might be interpreted as a prosodic break. However, research suggests that listeners' perceptions of prosody are not influenced by just the bottom-up acoustic signal. Instead, prosodic perception is at least partly driven by what the listener expects to hear given the syntactic, discourse, and pragmatic context. Recently, Buxó-Lugo and Watson (2016) investigated the perception of prosodic boundaries in contexts in which intonational boundaries are likely to occur versus perception of boundaries in locations where they are less likely. They presented participants with sentences like the ones below:

(3a) Put the big bowl on the tray.
(3b) Put the bowl that's big on the tray.

They found that participants were more likely to report hearing a boundary after “big” in (3b) compared to (3a), even when the acoustic information was digitally altered to be identical across contexts. This was true even when there was almost no acoustic evidence of a prosodic boundary. Buxó-Lugo and Watson (2016) argue that syntactic expectations drive this effect: listeners expect a boundary after “big” in (3b) because it coincides with a larger syntactic boundary than in (3a). In a corpus analysis, Cole, Mo, and Baek (2010) similarly found that the largest predictor of the presence of a boundary by naïve raters was syntactic context, followed by vowel duration. Consistent with these effects, Bishop (2012) found that the perception of prosodic prominence on a word can be manipulated by changing the discourse context in which that word appeared. Contexts in which the targets were discourse focused elicited higher prominence ratings compared to contexts in which the same target word was nondiscourse focused.

Together, these data suggest that prosody is not just a product of the stimulus. Instead, predictions based on our knowledge of the language influence how we perceive prosody, and evidence suggests these expectations can change somewhat quickly (Buxó-Lugo, 2017; Kurumada, Brown, & Tanenhaus, 2012; Roettger & Franke, 2019). While effects of expectations have been found for other areas of language processing (e.g., syntax, phonology), prosody is different in that there is little consensus over the structures and categories that make up prosodic representations (Wagner, McClay, & Mak, 2013 for discussion on potential meanings of the Rise-Fall-Rise intonation contour; for a more general discussion, see Ladd, 2014). Consequently, whereas in the case of syntax and phonemes one can more easily isolate the linguistic structure from the effect of expectations, in prosody effects of expectations might be confused with the linguistic structures themselves. This makes it difficult to determine what the grammatical structure of prosody actually is. Crucially, studying the effect of expectations on prosodic comprehension informs us about the relationship between listener's knowledge of a language and the acoustic signal they are trying to parse. For example, in Buxó-Lugo and Watson (2016), listeners report boundaries in unexpected locations, but only when the acoustic evidence was very strong. The weaker the acoustic evidence, the higher the likelihood that a listener would answer according to their expectations. This tells us that listeners balance acoustic information with expectations in ways that can be explicitly measured. Understanding these dynamics may be critical for understanding how prosody works.

4.2 | Prosody and communicative context

In English, acoustic prominence, or pitch accenting is used to signal that a referent is new to the discourse and/or potentially contrastive with something in the environment (e.g., the GREEN ornament, not the BLUE ornament). Although there is
general agreement that pitch, duration, and intensity are related to acoustic prominence (Breen et al., 2010), which of these combinations of features is most important is controversial. In fact, different labs have found different features to be important for signaling prominence (Beckman, 1986; Eady et al., 1986; Kochanski et al., 2005; Ladd & Morton, 1997; Sluijter & van Heuven, 1996). This may stem from the assumption that there is an idealized representation of prosody that generalizes across contexts. However, there is data suggesting that nonlinguistic factors such as speaker engagement and communicative context can influence how prosody is acoustically realized.

Buxó-Lugo, Toscano, and Watson (2018) compared two tasks that varied in speaker engagement: naming colors versus playing an interactive computer game with a partner. In both tasks, participants described three color patches in succession as in (4):

(4a) Green Blue Black.
(4b) Green Pink Black.

We varied the discourse status of the underlined word: it was either new, contrasted with the previous three-word description, or repeated. In the less engaging task, speakers described these color patches alone in a room without an interlocutor. In the more engaging task, participants played an online three-dimensional (3D) game with a partner. The participants were tasked with solving puzzles that allowed them to navigate an online dungeon. In critical trials, speakers had to describe color patches to their partner so that their partner could enter the colors into a combination lock that would allow them to proceed to the next trial. Clearly, the communicative or interpersonal stakes across the two types of trials differ enormously, which was reflected in speakers’ prosodic choices. Although speakers in both tasks differentiated discourse status using duration, it was only speakers in the more engaging task who used F0 to mark differences between given and new words.

While it may not be surprising that more engaged speakers produce more distinctive prosody, the idea that the communicative context might be important is not reflected in how prosody is typically studied. Although there is evidence that even speakers participating in nonengaging tasks can reliably produce intonational patterns that convey a specific meaning (e.g., incredulity; Goodhue, Harrison, Su, & Wagner, 2016), the effect we see in the above study might explain why different labs find different acoustic correlates across different tasks. Perhaps a lesson for language researchers is that anyone attempting to make judgments about the generic value of a sample of speech for understanding prosody must consider the degree to which a person is engaged in the communication task. Rather than assuming that prosodic structure is realized in the same way across contexts, researchers must acknowledge that prosody is realized in different ways depending on the goals of the speaker and the situations in which she is communicating. Critically, these data suggest that a theory of how prosody works must incorporate how context influences prosodic choices.

4.3 | Prosody and the production system

As we discussed above, there is a clear relationship between discourse status and acoustic prominence. New, unpredictable words are more likely to be prominent than given, predictable words. However, there is a great deal of evidence suggesting that low-level phonological encoding mechanisms may partly mediate this link.

Speakers are slower to produce words that share phonological structure with words that have already been produced (Watson, Buxó-Lugo, & Simmons, 2015; Yiu & Watson, 2015). For example, consider (5) below:

(5a) The beaker/house shrinks.
(5b) The beetle flashes.

In a scene description task from which the above stimuli were taken, Yiu and Watson (2015) found that speakers were more likely to lengthen the phonologically overlapping word “beetle” in (5b) when it was preceded by “beaker” as compared to a phonologically unrelated word like “house.” They argue that producing two similar sounding words creates interference during the phonological sequencing process that is linked to how quickly the sounds of those words are ordered.

In contrast to the phonological overlap results above, producing the same word twice leads to reduction (Jacobs, Yiu, Watson, & Dell, 2015; Kahn & Arnold, 2012; Lam & Watson, 2010, 2014), which may be the result of priming the phonological encoding system. Importantly, this ease comes not from retrieving the lexical item itself, but rather from sequencing the sounds in order, as speakers do not reduce words that they have only said in their heads to themselves (Jacobs et al., 2015). Evidence that repetition reduction effects are at least partly mediated by production processes comes from work by Jacobs
et al. (2015) showing that speakers also reduce repeated sequences of sounds that refer to an entirely different concept, such as homophones. That is, producing the word pie will make the subsequent production of pi shorter in duration (Jacobs et al., 2015).

Other work suggests that the semantic relatedness of words produced in succession might even impact production fluency, though the exact mechanisms are still debated (Balota, Boland, & Shields, 1989; Fink, Oppenheim, & Goldrick, 2018). Existing accounts that link aspects of speech timing to production ease (Arnold, 2008; Arnold, Kahn, & Pancani, 2012; Zerkle, Rosa, & Arnold, 2017) are also consistent with our perspective.

Ultimately, these data demonstrate that aspects of language that have traditionally been categorized as performance can have very real consequences for how prosody is realized in production and interpreted in comprehension. More importantly, the general finding that words that are easy to say are shorter than words that are more difficult to say poses a challenge for prosody researchers. In competence accounts, acoustic prominence, which is linked to duration, correlates with information structure (e.g., focus or discourse status). However, the data above suggests that words that are new to the discourse also happen to be words that are likely to be more difficult to produce, potentially conflating performance factors and linguistic cues. Understanding the many ways that prosodic production is directly and indirectly affected by other aspects of cognition may be important for uncovering system relationships between prosody and other aspects of linguistic structure.

5 | CONCLUSION

In sum, often-overlooked factors such as communicative context, top-down processing, and production pressures play an important role in the production and comprehension of prosody. Nongrammatical factors, such as an individual's boredom or engagement; the lexical, phonetic, and phonological properties of other entities in the discourse; and expectations about linguistic structure all impact prosodic processing. We have shown that speakers produce prosody in ways that do not necessarily reflect grammatical representations, but rather cognitive ones. Cognitive and grammatical representations may correlate in ways that make it difficult to disentangle the two in language production. Likewise, listeners are surprisingly sensitive to their own expectations of what the signal should sound like even when these expectations contradict the available acoustic evidence. We wish to make it clear that we are not arguing that competence-based approaches to prosody have not made contributions to our understanding of prosody, nor are we claiming that we should abandon these approaches completely. Instead we argue that understanding both prosodic knowledge and prosodic use in tandem will be critical for making progress in the study of prosody.

ACKNOWLEDGMENT

The first author was supported by the National Science Foundation under Grant No. 1557097. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

CONFLICT OF INTEREST

The authors have declared no conflicts of interest for this article.

AUTHOR CONTRIBUTIONS

Duane Watson, Cassandra Jacobs, and Andrés Buxó-Lugo: Conceptualization; writing-original draft, review, and editing.

ORCID

Duane G. Watson https://orcid.org/0000-0001-7191-4238

RELATED WIREs ARTICLES

Phonological regularity, perceptual biases, and the role of phonotactics in speech error analysis
Cognitive and linguistic biases in morphology learning
Variation, race, and multiracial identity in linguistic research
REFERENCES


---

**How to cite this article:** Watson DG, Jacobs CL, Buxó-Lugo A. Prosody indexes both competence and performance. *WIREs Cogn Sci*. 2020;11:e1522. https://doi.org/10.1002/wcs.1522