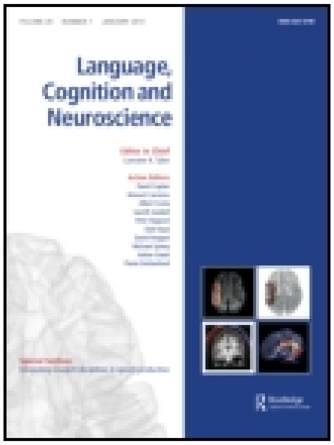

Language, Cognition and Neuroscience
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/plcp21

Prosody in context: a review

Jennifer Cole[a]

[a] Department of Linguistics, University of Illinois, 4080 Foreign Languages Building, 707 South Mathews, Urbana, IL 61801, USA

Published online: 08 Oct 2014.

To cite this article: Jennifer Cole (2015) Prosody in context: a review, Language, Cognition and Neuroscience, 30:1-2, 1-31, DOI: 10.1080/23273798.2014.963130

To link to this article: http://dx.doi.org/10.1080/23273798.2014.963130

Language, Cognition and Neuroscience, 2015
Vol. 30, Nos. 1–2, 1–31, http://dx.doi.org/10.1080/23273798.2014.963130

Prosody in context: a review

Jennifer Cole*

Department of Linguistics, University of Illinois, 4080 Foreign Languages Building, 707 South Mathews, Urbana, IL 61801, USA

(*Received 2 October 2012; accepted 6 June 2014*)

Prosody conveys information about the linguistic context of an utterance at every level of linguistic organisation, from the word up to the discourse context. Acoustic correlates of prosody cue this rich contextual information, but interpreting prosodic cues in terms of the lexical, syntactic and discourse information they encode also requires recognising prosodic variation due to speaker, language variety, speech style and other properties of the situational context. This review reveals the complex interaction among contextual factors that influence the phonological form and phonetic expression of prosody. Empirical challenges in prosodic transcription are discussed along with production evidence that reveals striking variability in the phonological encoding of prosody and in its phonetic expression. The review points to the need for a model of prosody that is robust to contextually driven variation affecting the production and perception of prosodic form.

Keywords: prosodic variation; intonation; prominence; prosodic phrasing; discourse context

1. Context, prosodic form and prosodic meaning

The prosodic form of a linguistic expression is linked in many ways to the context in which it is communicated. Prosodic form is influenced by context within the utterance and by the broader context of the discourse and situational setting of the utterance. Prosody also communicates information about those contexts. This review article considers the relationship between prosody and context, highlighting the role of prosody in conveying meaning related to the grammatical structure of words and their syntactic configuration (within-utterance context), and in conveying pragmatic meaning related to the broader linguistic context of the discourse and of the situational context, including speaker and addressee attributes. Also examined is the influence of linguistic context on the phonetic expression of prosody.

This review builds on Wagner and Watson's (2010) review article on prosody, which highlights prosodic prominence and boundaries in language processing. Here, we focus on how the same prosodic elements (prominence and boundaries) encode information about the linguistic and situational context. The dependencies between prosody and context discussed below highlight the role of prosody in the grammatical system of a language, its function in conveying discourse meaning and in managing interaction in dialogue. Regrettably, some topics relevant to an understanding of prosody in relation to context are only briefly discussed, e.g., the role of prosody in the expression of emotion, and prosodic imitation or entrainment in interactive dialogue, while others are not touched upon at all, e.g., the use of prosody in the performance of reported speech (Klewitz & Couper-Kuhlen, 1999) or quoting (Couper-Kuhlen, 1996), or prosody effects on code switching in a multilingual discourse context (Shenk, 2006).

We start in Section 2 by establishing the phonological basis of prosody and its phonetic expression. Section 3 discusses the role of prosody as a cue to the structural context at the lexical, syntactic and discourse levels, and Section 4 discusses the prosodic encoding of pragmatic meaning related to the discourse context. Section 5 makes the case for prosodic relativity in the need to evaluate prosodic features relative to the phonological-prosodic context in order to interpret those features as cues to the linguistic context at syntax and discourse levels, and to the situational context. In Section 6, we broaden our view to consider how prosody signals information about context having to do with the communicative situation.

2. Prosody in phonological form and its phonetic expression

Prosody is often described in terms of intonation and rhythm – the musical qualities of speech (Wennerstrom, 2001). Intonation and rhythm (or, more generally, timing) are *suprasegmental* aspects of speech because they define patterns that are largely independent of the segmental makeup (i.e., the consonant and vowel phones) of a given word or phrase. Suprasegmental properties relate to the auditory impression of pitch, loudness, and the duration and relative timing of phones, syllables and other speech units. These auditory qualities in turn depend on time-varying properties of the acoustic signal, including fundamental frequency (F0) and amplitude, and on the

*Email: jscole@illinois.edu

duration of acoustic intervals corresponding to phones and syllables.

The association of prosody with suprasegmentals has a long tradition in modern linguistics. For instance, Jakobson, Fant, and Halle (1951) distinguish the prosodic features of pitch, stress and duration from the 'inherent' distinctive features that characterise consonant and vowel sounds. For Jakobson, Fant, and Halle prosodic features are syntagmatic, encoding suprasegmental changes over position (or over time) within the utterance, such as an increase in pitch, loudness or syllable duration that marks a stressed syllable in English. On the other hand, 'inherent' (segmental) distinctive features are paradigmatic, marking oppositions among lexically contrastive consonants and vowels, such as, in English, the voicing distinction between /b/ and /p/, or the height distinction between /i/ and /e/. This functional distinction between prosodic and inherent features breaks down when we consider that in some languages suprasegmental features function lexically to distinguish words, as with tone in Chinese languages ([ma˥˥] 'mother' vs. [ma˧˥] 'hemp') or contrastive word stress in English (*IMport* vs. *imPORT*), or to mark grammatical features as with Spanish word stress (*HABlo* 'I speak' vs. *habLO* 's/he spoke'). Linguists often use the term '*intonation*' to refer to the systematic use of suprasegmental properties (or for some authors, just pitch) at the phrase or utterance level to mark linguistic information beyond word identity, i.e., post-lexical information (Cruttenden, 1986/1997; Gussenhoven, 2004; Ladd, 2008). For example, in connected speech, where words combine to form larger units such as phrases, utterances and discourse segments, languages may use pitch or other suprasegmental features to mark the beginning and/or final *boundaries* of such units, and in some languages, to mark the relative *prominence* (defined more precisely below) among words or larger constituents that occur in the same unit. These intonational features may function individually and in combination to convey utterance-level semantic and pragmatic meaning, as discussed in Sections 3 and 4.

A two-level view of prosody in terms of suprasegmentals at the word-level marking paradigmatic (lexical) contrast, and at the phrase-level conveying meaning about larger constituents (intonation), is appealing in its simplicity but still does not tell the whole story. The first problem concerns the function of prosody. Suprasegmental features at the word-level are not exclusively used to encode lexical contrast; in some languages they may be used syntagmatically and non-contrastively, paralleling their use in marking prosodic boundaries and prominence at the phrase level. Examples include languages with word-level stress at an invariant, fixed location (e.g., initial stress in Hungarian, see Varga, 1998), or languages that use tone or duration patterns to mark the beginning or end of a phonological word (e.g., word-final lengthening in English, see Beckman & Edwards, 1990). The second problem concerns the definition of prosody solely in terms of suprasegmentals. At the same locations where prosody is expressed through suprasegmental features we also often observe segmental effects, for example, on the acoustic parameters that encode voicing, manner or place of articulation. American English illustrates two such effects. Stop consonants show acoustic variation due to phrase-level prominence ('pitch accent') which affects the acoustic cues to place and voice features (Cole, Kim, Choi, & Hasegawa-Johnson, 2007), and vowels and sonorant consonants may exhibit glottalization when they occur initially in a prosodic phrase (Dilley, Shattuck-Hufnagel, & Ostendorf, 1996).

These various perspectives on prosody, in terms of a functional distinction (paradigmatic/syntagmatic), a level distinction (lexical/phrasal) or a featural distinction (suprasegmental/segmental), can be unified through reference to prosodic structure. Here, we may turn again to the musical metaphor: Just as there are tonal and temporal structures that give shape to the conventional melodies and rhythms of music, it is proposed that there are phonological structures that give rise to the prosody of spoken language. The idea is that a hierarchically organised phonological structure defines locations for the distribution of suprasegmental features (including, importantly, tone features), and the same structures influence the timing of phones and syllables, defining contexts for variation in their phonetic implementation (Beckman, 1996).[1]

Characterising prosody in terms of phonological structure affords prosodic features that have both paradigmatic and syntagmatic functions. Prosodic features can mark paradigmatic contrasts at the word level or above when two or more perceptually distinct features are licensed to occur in the same prosodic positions, where the choice among them signals different meanings.[2] In addition, prosodic features can function syntagmatically in demarcating the sequence of syllables, words, phrases, etc., when they occur at fixed locations in prosodic structure, such as the beginning or end of a prosodic word or phrase (Section 3).

Prosodic structure is defined both above and below the level of the phonological word. Sub-lexical structures are the syllable and (in at least some languages) the metrical foot, the latter serving to locate word-level stress (Hammond, 2011). Above the level of the word there are various proposals, but common structures include the prosodic phrase (possibly with two distinct levels), the utterance and the discourse segment as post-lexical prosodic constituents. There are differences among languages in the prosodic structures that have been proposed to account for observed phonological and phonetic patterns, e.g., in whether foot structure is posited within the prosodic phrase, or whether there is a level of structure below the prosodic phrase, such as the accentual phrase.

Notwithstanding these differences, most contemporary prosodic analyses characterise prosody in terms of *boundaries* that mark the edges of prosodic constituents such as words and phrases, and *prominence* (also stress, or accent) that is assigned to a designated element (the head) within the prosodic constituent at a given level. Phonological representations in this type of analysis are layered, with elements at a lower level combining to form constituents of a higher level (e.g., Nespor & Vogel, 1986/2007; Selkirk, 1984). For example, syllables combine to form stress feet, feet combine to form prosodic words, prosodic words combine to form prosodic phrases and so on up to the level of the utterance and above that, the discourse segment.

With prosody based in phonological structure, its expression in both segmental and suprasegmental properties can be understood as arising through two mechanisms. First, prosodic structure defines the locations where tone features are linked (e.g., at the edge of a phonological word, on a stressed (prominent) syllable within the word, or a phrase-final syllable), giving rise to the pitch contours that carry lexical, grammatical or pragmatic meaning. Second, prosodic structure influences the timing and magnitude of articulatory gestures for consonants and vowels. Generally speaking, gestures are lengthened and strengthened in certain prosodic positions (e.g., phrase-initially and in phrase-level prominent positions), while they are shortened and reduced in other positions (phrase-medially and in non-prominent positions), as shown in numerous works on English, including Edwards, Beckman, and Fletcher (1991); de Jong, Beckman, and Edwards (1993); Byrd and Saltzman (1998); Cho (2005); Byrd, Krivokapić, and Lee (2006); Cho and Keating (2009); Krivokapić and Byrd (2012) and in other languages (e.g., see works cited in Cho and Keating (2009) on prosodic strengthening and lengthening in Dutch, French, German, Korean, Spanish and other languages). These effects of prosodic context on articulation give rise to prosodically conditioned acoustic variation in consonants and vowels (i.e., segmental effects), as reported in numerous studies based on measures of acoustic duration, vowel formants, voice onset time (VOT), intensity and spectral measures of consonant place (e.g., Arabic: de Jong & Zawaydeh, 1999; American English: Cho, 2005; Cole et al., 2007; Turk & Shattuck-Hufnagel, 2007; Dutch: Cho & McQueen, 2005; Spanish: Ortega-Llebaria & Prieto, 2011; and many other works).

Figure 1 illustrates the prosodic structure assigned to an utterance of English, based on the autosegmental-metrical model (Ladd, 2008). The Intonational Phrase is the highest level of structure in this diagram, and the lowest level is the syllable. Elements designated as prominent are marked as 'strong' (subscripted *s*), and are the eligible anchor positions for pitch accents, represented in the diagram with the tone features H*, !H* and L*. In English, these pitch accents typically convey meaning related to the information status of a word or phrase. Also shown are tones that associate with the right edge of the intermediate prosodic phrase (H-, L-), and the intonational phrase (H%). The pitch accents and boundary tones together determine the pitch contour over the entire utterance.

We adopt the view of prosody as based in phonological structure in this review. Context effects on the distribution and realisation of prosodic features, such as pitch accents or boundary tones, are discussed in the following sections in terms of the mediating prosodic structures. We ask two general questions about the relationship between prosody (i.e., structure and associated features) and properties of the linguistic and situational context:

- What aspects of prosody are affected by context?
- What kinds of contextual information are conveyed through prosody?

In the following sections categorical prosodic features (structural, or features associated to prosodic structure) are distinguished from their phonetic expression in continuous-valued acoustic or articulatory parameters. Prosody in

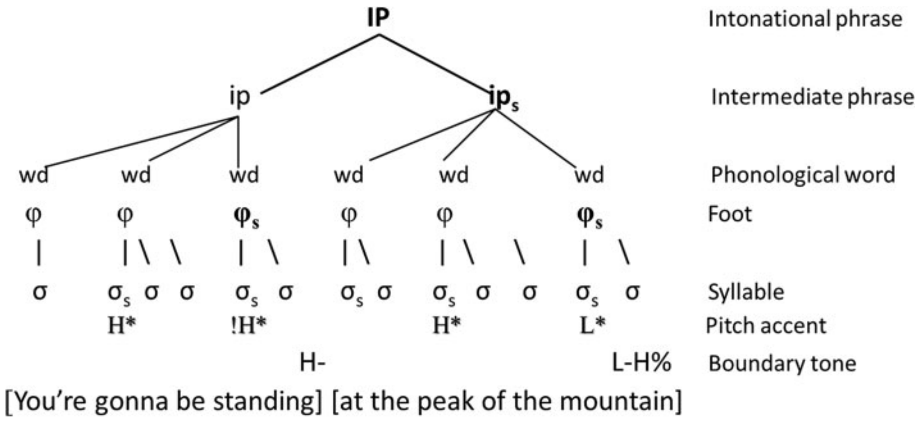

Figure 1. Diagram of an utterance showing hierarchically layered prosodic structure at the syllable, foot, word and phrase levels. Prominent positions are designated as strong (subscript 's'), and tone features are as described in the text.

the speech signal is discussed in terms of acoustic correlates of phonological prosodic structures and their associated tone features, or in terms of the listener's perception of those elements.

3. Prosody signals structure

With the understanding that prosodic structures license intonational (tone) features and influence the timing and magnitude of articulatory gestures, as described above, then it follows that patterns of variation in pitch, segment duration, voice quality and acoustic correlates of segment strength-of-articulation serve to index those prosodic structures. Furthermore, as discussed in this section, prosodic structures align (probabilistically) with lexical, syntactic and discourse structures, so it follows that the same patterns of segmental and suprasegmental variation also index the lexical, syntactic and discourse structures of the utterance. In other words, *the phonetic correlates of prosodic structure and associated prosodic features provide cues in the speech signal for the multi-layered linguistic context of words, phrases, utterances and larger discourse segments.* Referring back to the utterance diagrammed in Figure 1, in the spoken realisation of this utterance we expect to find a pitch plateau on the word *standing* expressing the high tone (H-) marking the intermediate phrase boundary, and a pitch rise from a low value at the start of the word *mountain* expressing the rising tone sequence (L-H%) marking the intonational boundary. These pitch patterns mark not only the ends of prosodic phrases, but due to the alignment of the prosodic phrases at their right edge with major syntactic boundaries, they also mark locations of syntactic juncture (in this example, between a verb and its PP adjunct), clause ending and possibly also the end of a discourse constituent.

This section reviews three kinds of linguistic structure that are cued by prosody and the phonetic expression of those prosodic cues. We want to know what kinds of structure from the linguistic context can be signalled by prosody, and whether there are common patterns in the prosodic marking of lexical, syntactic or discourse structure across languages. Here we consolidate findings from numerous studies that demonstrate the prosodic marking of linguistic constituent boundaries, starting from the word level and going up to the level of the discourse unit.

3.1. Prosody signals word segmentation

Many languages, including Arabic, English, Greek, Spanish, Turkish and a host of others, assign prosodic structure in the domain of the phonological word, locating stress-prominence on one syllable (or mora), typically at or near the left or right edge of the word. Stress prominence is culminative, in that there is a single primary stress for each phonological word (Hayes, 1995, p. 24), and it is delimitative, in that it signals the (precise or near) location of a word edge (Kager, 1995). Given these properties, the phonetic exponents of stress signal word units and provide cues that help listeners segment words from continuous speech (Finnish and Dutch: Vroomen, Tuomainen, & de Gelder, 1998; British English: Mattys, White, & Melhorn, 2005).

The phonetic implementation of word stress can be measured in acoustics and in articulation, and include segmental and suprasegmental effects (Gordon, 2011). The acoustic correlates of word stress differ among stress languages, but a common pattern is for stressed syllables to have greater magnitude than unstressed syllables in one or more acoustic dimensions. Early work on the acoustic correlates of primary word stress in English shows that in comparison with unstressed syllables, stressed syllables have greater duration, greater intensity and formant values showing vowels that are more peripheral in the vowel space (Fry, 1955; Lehiste, 1970; Lieberman, 1960). Fry also includes F0 as a primary correlate of word stress, but F0 in English is more likely a correlate of phrasal prominence and its associated pitch accent (Bolinger, 1958; Ladd, 2008, pp. 50–52; Pierrehumbert, 1980). Work by Sluijter and van Heuven (1996a) on Dutch shows that stressed syllables exhibit a change in spectral balance relative to unstressed syllables, which the authors interpret as indicating greater vocal effort in the production of stress. Sluijter and van Heuven (1996b) report similar findings for English (but cf. Campell & Beckman, 1997), as do Plag, Kunter, and Schramm (2011) in a study designed to differentiate between effects due to word-level stress and effects due to pitch accent (i.e., phrasal prominence). Ortega-Llebaria and Prieto (2011) look for acoustic correlates of stress in Central Catalan and Castilian Spanish, also differentiating stress effects from phrasal prominence and find duration to be the most consistent correlate, with little effects of word-level stress on overall intensity. This work also finds effects of stress on spectral balance (termed spectral 'tilt'), as was reported for Dutch and English, but only in Catalan, where the effects can be attributed to the pattern of vowel centralization in unstressed syllables. With Campbell and Beckman (1997), Ortega-Llebaria and Prieto suggest that vowel centralization may underlie the Dutch and English findings as well.

Looking at acoustic correlates of stress in other languages, we find differences among languages in the effect of stress on vowel centralization and overall intensity, but increased duration appears as a consistent correlate of stress across languages, when stress is considered independently of phrasal prominence (Arabic: de Jong & Zawaydeh, 1999; German: Dogil & Williams, 1999; Dutch: Cho & McQueen, 2005; see Ortega-Llebaria & Prieto, 2011, for other examples). When a stressed syllable has phrasal prominence, F0 effects are also frequently reported, again, most likely due to pitch accents.

Many languages exhibit effects of stress on the strength of consonant and vowel articulations, with strengthening in stressed syllables and weakening in unstressed syllables. Most generally, stress-induced strengthening is manifest in articulatory gestures that are longer in duration relative to the weakened gestures in unstressed syllables. Strengthened consonants also typically exhibit greater constriction, while strengthened vowels tend to be more peripheral in the vowel space, though other patterns of vowel strengthening are also reported (see discussion in Gordon, 2011). Building on Lindblom's (1990) idea of a continuum of articulation, de Jong (1995) interprets the phonetic correlates of stress in English as resulting from localised hyperarticulation of the stressed syllable. Findings from articulatory studies in a number of languages generally support this view (American English: Beckman & Edwards, 1994; de Jong et al., 1993 German: Mooshammer, Bombien, & Krivokapic, 2013; Mooshammer & Fuchs, 2002; Italian: Avesani, Vayra, & Zmarich, 2007).

The above examples show that across languages, word-level stress generally conditions greater overall intensity and duration as acoustic correlates, and strengthened segmental articulations. Duration in particular appears to be the most reliable correlate, signalling a local change in tempo that listeners can and do use, at least in some languages, as cues to word boundaries and aids to word segmentation in continuous speech recognition.

3.2. Prosody signals syntactic phrase boundaries

Words in continuous speech are grouped into prosodic phrases which may vary in length depending on speech rate, the degree of emphasis over the utterance, the syntactic structure of the utterance and other factors. Many studies have examined the correspondence between prosodic and syntactic phrase structures in English and in a host of other languages.[3] The general finding for English, for example, is that while there is a strong tendency for prosodic phrase edges to coincide with the edges of syntactic constituents, other constraints may intervene, resulting in mismatches between prosodic and syntactic phrases. An important constraint is that words that are linked through a meaning dependency preferentially occur in the same prosodic phrase: boundaries signal the relative independence of the upcoming words to the immediately preceding words (Breen, Watson, & Gibson, 2010; Frazier, Clifton, & Carlson, 2004; Selkirk, 1984).

The example utterances in (1), which are from two spontaneous speech corpora of American English, illustrate prosody–syntax alignment (1a), and mismatches (1b, c), with slashes marking lower (/) and higher levels of (//) prosodic boundary. The mismatches – for the prosodic boundaries after *to* in (1b) and after *wheel* in (1c) – depend, of course, on the syntactic analyses one assumes. Some researchers may view cases such as these as evidence for syntactic structures with displaced constituents such that prosodic boundaries may well align with syntactic structures at some level of analysis. The more general point here is that there is a relationship between prosodic and syntactic structures, but complexity arises due to complexity in one or both types of structure, and through the (real or apparent) misalignment of structures:[4]

(1) a. Now you're gonna go straight up / to the right of the abandoned cottage // and make a left
b. More 'n more people // y'know // struggling to / get a gun and / shoot someone
c. So you're gonna go / between / the mill wheel // and the mountain

The examples in (1) distinguish two levels of prosodic phrasing, as does the prosodic structure diagrammed in Figure 1 above. The autosegmental-metrical model of intonation posits this distinction in English on the basis of the F0 configurations observed preceding a phrasal juncture (Beckman & Pierrehumbert, 1986; Ladd, 2008, pp. 101–104; Pierrehumbert, 1980). More complex contours are observed at the end of phrases that are also judged to have stronger juncture, e.g., at the end of an utterance. This difference is modelled by allowing a single tone feature, termed the 'phrase accent' (annotated H- or L-) to mark a lower-level phrase boundary (the intermediate phrase in Figure 1), with an additional tone (H% or L%) appended at the end of a higher-level boundary (the intonational phrase in Figure 1). Combinations of phrase accent and boundary tone can result in tonal contours, such as a rising L-H%.

Experimental findings that point to the alignment of prosodic and syntactic phrase edges, including those cited above, are largely based on read speech, with sentences designed by the experimenter to test factors that influence prosodic phrasing. Finding evidence of a prosody–syntax alignment in spontaneous speech is more challenging due to (1) the difficulty in assigning syntactic structures to spontaneous speech given the prevalence of sentence fragments, run-on sentences and disfluency; and (2) the lack of control over the myriad factors other than syntactic context that may also influence prosodic phrasing. Schafer, Speer, Warren, and White (2000) approach the problem through the use of a cooperative game task that effectively constrains the range of lexical items, syntactic structures and pragmatic context while still eliciting spontaneous speech from their study participants. They report that speakers use prosodic structure to signal syntactic structure in disambiguating early vs. late closure, locating a stronger prosodic boundary (//) at the location of the higher syntactic boundary in utterances such as

'When that moves the square // it should land in a good spot', and 'When that moves // the square will encounter a cookie'.

While not every syntactic juncture coincides with a prosodic phrase juncture, prosodic boundaries perceived by transcribers tend to align with syntactic boundaries. Two corpus studies examine prosody–syntax alignment in unrestricted, conversational speech which has been annotated for its syntactic structures using automatic methods, or by trained experts.[5] Calhoun (2006) reports that fully 72.3% of (syntactic) clause boundaries coincide with prosodic phrase boundaries, and she further shows that a statistical model that incorporates syntactic, semantic and acoustic features can predict over 87% of prosodic phrase boundaries. In a study with similar speech materials but with prosody annotations collected from groups of untrained, naive transcribers, Cole, Mo, and Baek (2010) report that 51% of clause-final syntactic boundaries ('S' or 'S-bar') are judged as locations of prosodic boundaries by over half the transcribers (15–22 total), with boundaries perceived less consistently at lower-level syntactic junctures (e.g., NP, PP); dropping the threshold to count locations where one or more transcribers perceive a prosodic boundary results in prosodic boundaries at 98% of the clause boundaries in the corpus.

Prosodic phrases may be phonetically marked at their left and right boundaries, and here we cite a handful of studies from a larger body of work on acoustic and articulatory correlates of prosodic phrase boundaries. It bears noting, again, that most of the works cited below are based on read speech materials, which is significant in light of the claim that prosodic phrasing is influenced by factors related to speech planning (Watson & Gibson, 2004). It is evident even to the casual observer that there are differences between read and spontaneous speech, arguably due to differences in speech planning, and these differences extend to prosody (Blaauw, 1994). Further research is needed to establish whether there are differences in the phonetic cues to prosodic boundaries in read vs. spontaneous speech, and to compare phonetic cues to boundaries across languages that differ in their syntactic or prosodic structures.

Prosodic phrasing is phonetically realised through the location of silent pause at prosodic phrase junctures and through lengthening of segments preceding a prosodic boundary (final lengthening). Both are well-studied phenomena by researchers in phonetics, spoken language processing and speech technologies.[6] Acoustic evidence for final lengthening is shown for American English (Cole, Mo, & Baek, 2010; Wightman, Shattuck-Hufnagel, Ostendorf, & Price, 1992; Yoon, Cole, & Hasegawa-Johnson, 2007), where lengthening may extend as far back as the primary stressed syllable in antepenultimate position (Turk & Shattuck-Hufnagel, 2007). Acoustic evidence of final lengthening is also reported for other varieties of English (British English: Hirst & Bouzon, 2005; Singapore English: Low, Grabe, & Nolan, 2001), and for many other languages (e.g., Mandarin Chinese: Cao, 2004; Dutch: Cambier-Langeveld, 1997; Cho & McQueen, 2005; Hungarian: Hockey & Fagyal, 1999; Russian: Volskaya & Stepanova, 2004; Swedish: Horne, Strangert, & Heldner, 1995). Articulatory studies with American English show that lengthening effects are gradient, decreasing with distance from the boundary (Byrd, Kaun, Narayanan, & Saltzman, 2000; Byrd & Saltzman, 1998; Edwards et al., 1991; Krivokapić & Byrd, 2012). Prosodic phrase boundaries are also phonetically marked by lengthening and articulatory strengthening of the segment immediately following a boundary (initial strengthening), with acoustic and articulatory evidence from English (Fougeron & Keating, 1997; Keating, Cho, Fougeron, & Hsu, 2003) and from other languages (Dutch: Cho & McQueen, 2005; French: Fougeron, 2001; Korean: Cho & Keating, 2001; see also Keating et al., 2003). Comparing initial strengthening in Dutch and English, Cho and McQueen (2005) argue that strengthening effects may be language-specific, affecting those acoustic parameters that cue phonological contrast in the language.

Prosodic phrasing is also phonetically signalled through F0 patterns that extend over material preceding the right-edge boundary and in some languages with F0 marking at the left-edge boundary. These F0 patterns realise the edge tones associated with the prosodic boundary. The intonational encoding of prosodic phrase boundaries has been a primary focus of research for many decades, and is addressed in every framework proposed for the analysis of intonation (representative works include Bolinger, 1982; Halliday, 1967; Hirst, Di Cristo, & Espesser, 2000; Pierrehumbert, 1980; t'Hart, Collier, & Cohen, 1990). Different F0 contours, representing distinct configurations of edge tones, are associated with differences in the pragmatic meaning of the utterance, as discussed in Section 4.2. Here we note only that the F0 contours that mark the end of a prosodic phrase are distinct from F0 contours that express pitch accents primarily in *not* being prominence-lending, which is to say that they extend over one or more final syllables regardless of the status of those syllables as prominent (i.e., as bearing word-level stress or phrasal prominence).

In additional to the lengthening, strengthening and F0 effects of prosodic phrase boundaries, certain laryngeal effects such as glottalization of domain-initial sonorants and creaky voice in the final region of the prosodic phrase are observed in American English (Dilley et al., 1996; Redi & Shattuck-Hufnagel, 2001). These phonetic effects of boundaries appear not to be as common cross-linguistically as lengthening, strengthening and F0 effects.

The studies cited above, and many others, establish that across languages prosodic phrase structure is acoustically signalled through pause and through increased phone and syllable duration in the vicinity of a prosodic boundary, and through the presence of non-prominence-lending F0 contours that realise edge tones. Voice quality effects and acoustic effects of articulatory strengthening may provide additional boundary cues in some languages. Because prosodic phrase boundaries tend to align with syntactic phrase edges, prosodic boundary cues are to some degree indicative of the syntactic context that coincides with the prosodic boundary. Indeed, a number of perception studies show that listeners are sensitive to prosodic boundary cues and use prosodic information in interpreting sentence structure (American English: Carlson, Clifton, & Frazier, 2001; Clifton, Carlson, & Frazier, 2002; Schafer et al., 2000; Snedeker & Trueswell, 2003).

Assessing the role of prosody in cueing syntactic structure in conversational speech is less straightforward than in read speech. The reliability of prosodic cues to syntactic boundaries in conversational, unscripted speech is limited by the fact that, at least for English, prosodic boundaries may be optional, both clause-internally, and at locations of clause-level syntactic boundaries. For example, Snedeker and Trueswell (2003) show that speakers in their study were not consistent in producing prosodic boundaries in contexts where such boundaries can disambiguate competing syntactic analyses, e.g., in distinguishing PP attachment in NP modifier readings [*Tap (the frog with the flower*)] vs. instrumental readings [*Tap* (*the frog*) (*with the flower*)], even when speakers were aware of the ambiguity and were engaged in interactive, instruction-giving dialogue. From this study it appears that while the presence of a boundary is often informative, the absence of a prosodic boundary at a word edge is not a reliable cue to the immediate syntactic context of the word. On the other hand, Kraljic and Brennan (2005) find consistent prosodic marking of syntactic structure in spontaneous English speech in tasks where speakers are engaged in interactive communication, and they suggest that the failure to prosodically encode syntactic boundaries occurs in non-interactive or routine speaking tasks (as in the Snedeker & Trueswell study). Another complicating factor for the mapping from prosodic boundary to syntactic boundary in speech comprehension is the distinction between lower- and higher-level prosodic boundaries. Although there is a tendency for speakers to use higher prosodic boundaries at higher syntactic junctures (Ladd, 2008, pp. 288–299; Wagner, 2005, 2010), there is not a strict association between a specific level or type of syntactic juncture and the occurrence of either an intermediate or intonational phrase boundary. Rather, what is important is the strength of a prosodic boundary *relative to* other nearby prosodic boundaries, which should correspond to the relative strength of the corresponding syntactic boundaries at the same locations (Clifton et al., 2002; Schafer et al., 2000; Wagner, 2005, 2010). These and other examples of prosodic relativity are discussed in Section 5.

Evidence from prosodic transcription studies points to the difficulties listeners face in interpreting prosodic cues to syntactic structure. For example, trained transcribers for American English show high agreement rates for the location of prosodic boundaries, with much lower rates for discrimination between intermediate vs. intonational phrase boundaries (Pitrelli, Beckman, & Hirschberg, 1994; Yoon, Chavarría, Cole, & Hasegawa-Johnson, 2004), indicating that the acoustic cues (or other criteria) for differentiating boundary level are not always very clear. The difficulty in perceiving distinctions in boundary strength is surprising in light of evidence from production studies with read speech showing that speakers do, in some instances, produce distinctions in prosodic boundary strength in structurally ambiguous sentences, where the stronger prosodic boundary signals the higher syntactic boundary (see further discussion in Section 5.3). Nonetheless, the difficulty transcribers have in perceiving boundary-level distinctions suggests that speakers are not consistent in producing boundary-level distinctions. At the minimum, we acknowledge that uncertainty in the perception of boundary strength reduces the reliability, and therefore also the validity of the prosodic boundary as a cue to syntactic structure.

Other factors that impact the identification of syntactic boundaries on the basis of prosodic boundaries relate to speaker variability in the selection of acoustic cues that encode boundaries, and ambiguity in the interpretation of boundary cues. For example, in their study of prosody perception and production in Southern British English, Peppé, Wells, and Maxim (2000) report that speakers vary in their use of pause, lengthening and F0 contours to mark an utterance-internal prosodic boundary in producing lists with two elements, '[X Y] and Z', or three elements, 'X, Y, and Z'. Another factor is the overlap in the acoustic cues for prosodic boundaries and the acoustic expression of disfluency (Shriberg, 2001). For instance, silent pause occurs frequently in disfluencies that express hesitation, or preceding a disfluency repair (Nakatani & Hirschberg, 1994). Disfluent pauses tend to occur early in an utterance, and need not coincide with major syntactic juncture, yet they may be difficult to distinguish on phonetic grounds from pauses that mark fluent prosodic and syntactic boundaries (Cole et al., 2005). Yet another limiting factor is the fact that some acoustic effects (increased duration, some F0 contours) are potentially ambiguous as cues to prosodic boundary or as cues to phrasal prominence, since they arise variously under both conditions. We return in Section 7.1 to the question of the interaction among contextual factors that affect the realisation of prosody, e.g., on F0 as a cue to both pitch accent and prosodic boundary.

At present we must acknowledge that despite the fact that researchers find acoustic cues to prosodic boundary, our understanding of how these cues function in signalling information about the syntactic context in the vicinity of a cue is limited by (1) the inconsistent relationship between prosodic and syntactic structure, and (2) variability in selection of acoustic cues that encode prosodic boundaries and (3) the use of potentially ambiguous acoustic cues to prosodic boundaries.

3.3. Prosody signals discourse structure and rhetorical structure

Similar to the manner in which prosody provides cues to syntactic boundaries, there is also prosodic marking of the boundaries of discourse segments in interactive speech. Prosodic encoding of discourse structure in particular has been investigated in the context of research in conversational interaction, and for the development of computer dialogue processing systems. Commonly observed patterns in conversational and read speech are high F0 and increased F0 range in the initial phrase of a discourse unit (also topic unit, or rhetorical unit for read text),[7] with declining F0 across utterances within the unit, ending in a low F0 at the end of the discourse unit or turn (English: Grosz & Hirschberg, 1992; Herman, 2000; Menn & Boyce, 1982; Yule, 1980; Dutch: Geluykens & Swerts, 1994; Sluijter & Terken, 1993; Swerts, 1997; Swerts & Geluykens, 1993, 1994; Mandarin: Tseng, Pin, Lee, Wang, & Chen, 2005; Wang & Xu, 2011). Following the end of a topic unit (as a type of discourse unit), an F0 reset or increase is found at the start of a new topic (American English: Nakajima & Allen, 1993; Dutch: Geluykens & Swerts, 1994; Mandarin: Wang & Xu, 2011). A related finding is that speakers use a flat mid-level or high pitch at the end of an utterance to mark the continuation of a topic, where it signals that the speaker does not want to relinquish the floor to their conversation partner (English: Laskowski, Helder, & Edlund, 2009; Brazilian Portuguese: Oliveira & Freitas, 2008). Many of the studies just cited also report longer pauses at the end of a discourse unit than at the end of prosodic phrases that are medial in the discourse unit. Additional durational effects are observed in speech rate changes in the vicinity of a discourse boundary. For instance, Smith (2004) finds increased final lengthening and slower speech rate at the end of a textually defined discourse segment and into the initial breath group of the following discourse segment, while den Ouden, Noordman, and Terken (2009) find slower speech rate on the nucleus of the discourse segment (described as the central part of a text span), and faster speech rate across the juncture of text segments that are causally related (i.e., not topic shift junctures).

The beginning and end of a discourse (or rhetorical) segment are also, of course, the beginning and end of a prosodic phrase, but the discourse boundary is distinguished from a (discourse-medial) prosodic phrase boundary by the acoustic enhancement of the phrase-marking prosody. For instance, Geluykens and Swerts (1994), in a study of Dutch spontaneous instruction-giving dialogues, show that the lowest F0 values are observed at the end of a clause that is also the end of a topic unit, which is operationally defined in their task as the end of a step in the instruction sequence. In another paper from the same study, Swerts and Geluykens (1993) report that in addition to F0 marking of the edges of a discourse segment, speakers also produce the longest pauses at the end of a topic unit.

As with syntactic boundaries, the relationship between prosodic marking and discourse structure is not simple, and exhibits substantial variability. Speakers do not consistently produce prosodic cues that identify the beginning or end of a discourse unit, or cues that signal the end of the speaker's turn. For example, Caspers (2003), in a study of Dutch MapTask dialogues, finds that the end of a discourse segment is not always marked with intonational features, and many discourse-final locations are described as being marked with nothing more than typical 'sentence-final' intonation. Of course, there are many acoustic correlates of prosodic structure, so a failure to find evidence from one acoustic correlate for discourse boundaries, such as F0, does not mean that the discourse boundary is not prosodically marked. It is likely that listeners' ability to perceive discourse boundaries rests on the combined evidence from multiple acoustic cues. Evidence from perception studies supports this view. Swerts and Geluykens (1993, 1994) and Geluykens and Swerts (1994) show that Dutch listeners are sensitive to F0 and pause as prosodic cues to discourse structure marking the end-of-turn in Dutch and can detect such boundaries even in band-pass filtered speech that reveals little of the lexical content of the speech. In related work on Dutch, Swerts, Bouwhuis, and Collier (1994) present results from a perception study showing that the register, range and shape of the pitch contour at the end of an utterance convey the degree of finality, e.g., expressing the end of a discourse unit.

Herman (2000) tests American English listeners' sensitivity to discourse prosody with naturally produced, read speech, and finds that although speakers do not always produce prosodic cues that effectively distinguish discourse-final utterances (2b) from non-final utterances (2a), when such cues are present listeners use them to identify the discourse context of an utterance that is presented in isolation. Herman reports the presence of acoustic cues to discourse boundary in those utterances where listeners detected a boundary, in measures of F0, duration and root-mean-square (RMS) intensity over the entire utterance and in cues that are localised in the final syllable:

(2) Example dialogues from Herman (2000).

Prompt produced by experimenter:
Are you going to need to change advisors?

Response produced by participant:
I was talking to my advisor the other day.
She said she's going to be doing field research in Kenya.

(a) She'll be gone for the whole year.
(b) It's o.k. to have an advisor who's away, but only for one quarter.

Further evidence that listeners are sensitive to prosodic encoding of discourse boundaries comes from studies showing that listeners judge speech as more natural when it displays typical discourse prosody. Thus, Sluijter and Terken (1993) find that Dutch listeners judge diphone-synthesised speech as more natural when it displays paragraph intonation patterns with F0 declination such as were produced by Dutch speakers in their study. Similarly, Smith (2004) found that American English listeners judge resynthesized speech generated with final lengthening and pause duration patterns based on native speaker productions as more natural than the same speech resynthesized without F0 and pause as cues to discourse prosody.

To summarise, many studies show prosodic marking of discourse structure, with boundaries marked by shifts in F0, local segment duration and (in some studies) intensity in the vicinity of the discourse boundary, and by pauses at the juncture between discourse segments. The prosodic cues to discourse boundaries are largely the same as the cues that mark prosodic phrase boundaries, but in discourse boundary locations the cues are enhanced: more extreme F0 maxima and minima at the beginning and end of a discourse segment, respectively, increased final lengthening effects, and longer pauses. Many studies show variability in the prosodic encoding of discourse structure, both across experimental items and across speakers, so listeners cannot rely exclusively on such cues to identify discourse context; but when prosodic cues are available, listeners do appear to make use of them in detecting discourse boundaries and in evaluating the degree of discourse cohesion or juncture between successive among utterances in a discourse.

4. Prosody signals discourse meaning

One of the most salient aspects of prosody in English is its role in communicating the meaning of a phrase or sentence as it relates to the discourse context. The function of prosody in communicating those elements of meaning that lie ‘beyond words’ is widely recognised and has been an increasing focus of research in recent years (Barth-Weingarten, Dehé, & Wichmann, 2009; Wichmann, 2011). This section examines how prosody conveys the meaning of an utterance as it relates to the discourse context, as one component of pragmatic meaning (the other being meaning related to the situational context, discussed in Section 6). The function of prosody in conveying pragmatic meaning across languages is claimed to derive from deep-rooted biological mechanisms (Section 4.6), yet despite this common basis, there are sometimes remarkable differences among languages in the aspects of discourse meaning that are prosodically encoded, and in the phonetic expression of prosodic features marking discourse meaning. This section reviews three types of discourse meaning signalled by prosody – focus or information status, illocutionary force and affective meaning – and also discusses the role of prosody in managing discourse interactions between speakers. The status of prosodic encoding of discourse meaning is discussed in relation to its variability, its use by listeners and by considering its basis in biological mechanisms.

4.1. Prosodic prominence signals focus and information status

In many languages, prosodic prominence conveys information about the semantic and pragmatic context by signalling the status of a word or constituent as new to the discourse (so-called ‘broad’ or informational focus), or as having contrastive or narrow-scope focus (American English: Breen, Fedorenko, Wagner, & Gibson, 2010). For example, English and German signal focus and information status (i.e., new/given, or the accessibility of a word given the discourse context, see Wagner & Watson, 2010) through the assignment of phrasal prominence and an associated pitch accent on a focused or discourse-new word (e.g., English: Brown, 1983; Büring, 2006; Chafe, 1987; Ladd, 2008, pp. 213–259; Rooth, 1992; Selkirk, 1995; German: Fery̋, 1993; Fery̋ & Kügler, 2008).

In English, the association between pitch accents on one hand, and focus or information status on the other is mediated through phrase-level metrical structure (i.e., the strong–weak patterning of feet as in Figure 1). A word located in a strong position in the phrase-level metrical structure is *prominent* relative to a word in a weak (or hierarchically lower) position, and the rightmost prominent word in the phrase is the head or nucleus of the phrase – unless a preceding word has narrow or contrastive focus, in which case the focused word is assigned the nuclear prominence (Calhoun, 2010a, 2010b; Katz & Selkirk, 2011; Pierrehumbert, 1980; Selkirk, 1995;). The nucleus is assigned an obligatory pitch accent. Additional pitch accents are optionally assigned to prenuclear words (Bolinger, 1986) based on their informativeness (Calhoun, 2010b), and exhibit a tendency towards rhythmic alternation and early placement in the phrase (Calhoun, 2010a; Shattuck-Hufnagel, Ostendorf, & Ross, 1994). In English the nuclear prominence is the highest prominence in the phrase-level metrical structure, and is sometimes described as having the greatest (perceptual) salience (Ladd, 2008, pp. 131–147). For example, in ***John*** *ate the* ***APPLE***, both *John* and *apple* are prominent, but as the nuclear element, *apple* has greater prominence and an obligatory pitch accent. The nuclear prominence can occur earlier in the sentence when it marks an emphatic or contrastive focus, as in ***JOHN*** *ate the apple* uttered in response to the question *Did Mary eat the apple?*.

Non-final nuclear prominence in English can also occur under other conditions related to discourse meaning. Gussenhoven (1984) observes that nuclear prominence (*accent* in his terminology) is normally assigned to an argument in an argument–predicate construction that constitutes a single focus domain ([**A** P]), e.g., *Our* ***DOG****'s disappeared*, or *The* ***COFFEEMAKER*** *broke,* a pattern that Ladd (2008, p. 247) attributes to the fact that arguments typically carry more semantic weight than predicates. Accordingly, Ladd and Gussenhoven observe that nuclear prominence will occur on the predicate when its argument is a semantically light pronoun or indefinite (i.e., the argument is not lexically filled), as in *He* ***DISAPPEARED*** vs. *Our* ***DOG****'s disappeared,* or *I* ***ATE*** *something* vs. *I ate an* ***APPLE*** (Ladd, 2008, pp. 223–251). Gussenhoven (1984) argues that the status of the sentence as eventive (i.e., expressing an actual event) is another factor affecting nuclear prominence, with prominence assigned to the predicate rather than the argument in non-eventive sentences (i.e., hypothetical, conditional or definitional), as in *Apples are* ***FRUIT****, People will be* ***SHOT***, or *Dogs must be* ***CARRIED***.[8]

The description above reveals the complexity of the system that assigns prominence (via metrical structure) to words within a phrase in English (and similarly, in Dutch). Overall, there is a strong tendency for metrical structure to be assigned to an utterance such that a word that carries new information or a word with narrow focus will have nuclear prominence. If this were an absolute rule, it would result in a 1–1 mapping between information status/focus and prominence. But identifying information status or focus from prosody is challenging in English because information status or focus are not the only factors that determine metrical parsing and pitch accent placement. As shown by Calhoun (2010b), factors related to rhythmic structure and 'accentability' (i.e., predictability, referent accessibility) also play a role in the location of pitch accents. Calhoun offers the example in (3) from the Switchboard corpus of American English telephone conversation speech (Godfrey, Holliman, & McDaniel, 1992). The final utterance is of interest here. Although the noun phrase *the prison system* is focused (as the response to the question), and bears a pitch accent, the utterance is produced with additional pitch accents on *I* and *cut* (as identified in Calhoun's prosodic annotation and also visible from the pitch display shown there). These prenuclear pitch accents on words that are discourse-given are motivated by metrical structure and the preference in English for rhythmic alternation, which is attained in this utterance: ***I*** *would* ***cut*** *the **pris**on systems:*

(3) B: my first comments on the budget
A: what would be the first thing you'd cut? defense?
B: Surprisingly, no
B: **I** would **cut** [the **PRISON** systems]F

Calhoun (2010a) offers another example of a mismatch between information status and pitch accent from the Switchboard corpus in (4). The final utterance, produced as a single prosodic phrase (marked with parentheses), was produced by Speaker A with the nuclear pitch accent on a discourse-given word, *goods,* instead of locating it on *in* or *here,* both of which are discourse-new and more informative, and which are also closer to the right edge of the phrase – the default position for nuclear prominence in English:

(4) A: the deficit basically is the trade surplus … we have … more money going out, and too many goods coming into this country … like Japan still does not let us compete fairly in their country, and obviously the demand for their goods is quite high here so
A: [**they** can get their **GOODS** in **here]**

Calhoun explains the absence of a nuclear pitch accent on *in* or *here* as reflecting the relatively low accentability of these words:

> [w]e do not expect *in*, a preposition, to be accented; and so if it is, narrow scope is implied – for example, *in* as opposed to *out* of here – which is odd in the context. Similarly, an accent on *here* is not expected since it is semantically light, so could introduce an unintended alternative set of other places. Therefore, it [the pitch accent /JC] has to occur on *goods.* (p. 27; italics added here)

It follows that the post-nuclear pitch accent on *here* must be attributed to rhythm. The overall picture, then, is that while the presence of a salient pitch accent tends to be a mark of new information or focus, it is not always so, and the absence of pitch accent cannot always be interpreted as indicating that a word or constituent is discourse-given and not focused. Critical to the analysis of a given utterance is that the pitch accents be interpreted *relative to* other features of the prosodic structure in the utterance, and in light of the lexical and syntactic properties of the utterance, a point also emphasised by Katz and Selkirk (2011). We return to these points in Section 5.

It is important to note that pitch accents are not universally used to convey focus or information status.

For example, in Spanish (Hualde, 2000), pitch accents are assigned to (nearly) every content word, so the presence or absence of a pitch accent does not convey information about phrase-level distinctions in prominence or information status. Another example is French (Delais-Roussarie & Rialland, 2007), where pitch accent is associated with the final position in the phrase, and it therefore signals phrase structure rather than phrasal distinctions in prominence. Similarly, in Korean pitch accents mark prosodic phrase structure, and focus is expressed primarily through its effects on phrasing (Jun, 2005a). See Frota (2000, pp. 15–31) for a typology of prosodic focus-marking mechanisms across languages.

Focus and information status are phonetically expressed in American English in the acoustic correlates of prominence, including increased duration and intensity (Cole, Mo, & Hasegawa-Johnson, 2010), through F0 correlates of pitch accents associated with prominence (Breen, Fedorenko, et al., 2010; Cooper, Eady, & Mueller, 1985; Eady & Cooper, 1986; Eady, Cooper, Klouda, Mueller, & Lotts, 1986; Katz & Selkirk, 2011; Xu & Xu, 2005), and through acoustic effects on vowel formants that reflect local hyper-articulation and in other contrast-enhancing effects (Cho, 2005; de Jong, 1995).[9]

The shape of the F0 contour marking focus or discourse-newness is determined by the particular tone or tone sequence of the pitch accent, e.g., H* vs. L*+H (Beckman, Hirschberg, & Shattuck-Hufnagel, 2005; Gussenhoven, 1984; Pierrehumbert, 1980). A number of studies have investigated the pragmatic meaning associated with the different pitch accent types in English and Dutch. We discuss two here for illustration. Krahmer and Swerts (2001) compare intonation patterns in contrastive and non-contrastive focus in Dutch, and although they find evidence from production data for a difference in the distribution of pitch accents under the two focus conditions, they argue that these differences can be attributed to the status of the pitch accent as nuclear (phrase-final) or pre-nuclear, so the findings do not support the claim that the choice of pitch accent directly encodes focus type. Breen, Fedorenko, et al. (2010) come to a somewhat different conclusion in their study comparing focus type (contrastive vs. non-contrastive) and focus breadth (narrow vs. broad) in American English, where they find that these distinctions are marked in the phonetic correlates of the pitch accents that mark focus. Breen and colleagues examined pitch accent in utterances of the type in (5b) elicited in response to question prompts as in (5a). They compare acoustic measures across sentences with different focus conditions and their findings show that the distinction between contrastive and non-contrastive focus is marked by lower F0 (mean and maximum) and higher intensity on contrastive focus, and that in SVO sentences with nuclear prominence on the object, narrow focus (on the object) is distinguished from broad focus (on the VP or on the entire sentence) by higher F0 and longer duration for the object in the narrow focus condition, and in the wide focus condition by higher intensity and F0 and longer duration on prefocal words:

(5) a. prompting questions:
What happened last night? (broad focus)
Who fried an omelet? (subject focus)
What did Damon do to an omelet? (verb focus)
What did Damon fry? (object focus)
b. Damon fried an omelet.

Looking beyond English, there are many studies that examine the prosodic marking of focus and information status in other languages. Ladd (2008, ch. 6.2) discusses differences among languages in their sentence accent patterns as they relate to the marking of information status. He describes a salient difference between Romance and Germanic languages concerning the absence of accent (‘deaccenting’) on words that are discourse-given, which is quite common in English, but not standard in Italian or Romanian. Ladd’s observations are consistent with the findings of Swerts, Krahmer, and Avesani (2002) in a controlled study comparing Italian and Dutch for prosodic marking of information status (newness and contrastive focus).

Looking at the phonetic expression of prosodic features that mark information status, although the detailed patterns vary across languages, many studies report longer duration and/or F0 contours with expanded F0 range for focused or discourse-new words relative to words that precede or follow them in the utterance (e.g., for Arabic: de Jong & Zawaydeh, 2002; Dutch: Cambier-Langeveld & Turk, 1999; Cho & McQueen, 2005; German: Baumann & Grice, 2006; Baumann & Hadelich, 2003; Feŕy & Kügler, 2008; Mandarin Chinese: Xu, 1999; Xu & Wang, 2001; for other languages, see Jun, 2005b; Ladd, 2008, ch. 6.2).

The studies cited above look at evidence for prosodic marking of focus and information status in speakers’ productions. There have also been a number of studies showing that listeners perceive prominence in relation to F0 as a correlate of pitch accent, and to other acoustic cues to prominence, and use such cues to interpret the focus and information status of referring expressions. One general finding is that listeners tend to interpret unaccented words as given or accessible based on the prior discourse, while accented words are interpreted as new information, and identified with less accessible referents (Dutch: Terken & Nooteboom, 1987; German: Weber, Braun, & Crocker, 2006; British English: Chen, den Os, & de Ruiter, 2007; American English: Birch & Clifton, 1995; Dahan, Tanenhaus, & Chambers, 2002; Ito & Speer, 2008; but cf., Arnold, 2008, for evidence against the accented word bias). The interpretation of information status based on F0 appears not to rely on a specific F0 value (e.g., peak height) on a given word, but rather requires considering the F0 contours marking pitch accents in relation to one another and to the word's position as early or late in the utterance (Dutch: Rump & Collier, 1996). The study by Krahmer and Swerts' (2001) on contrastive accents in Dutch, discussed earlier, finds similar evidence for the perception of prominence differences in relation to focus (see further discussion in Section 5.4).

Given that the accent status of a word influences the interpretation of its information status, it follows that any mismatch between the prior discourse context and the accentuation of a word should disrupt discourse processing. This prediction is confirmed in several studies, including some cited in the previous paragraph, which show evidence from eye-tracking experiments that an initial assignment of referent to a word can be misguided by inappropriate accentuation (Arnold, 2008; Dahan et al., 2002; Ito & Speer, 2008), and with evidence from event-related potentials (ERPs) for difficulty in semantic integration with inappropriate focus-marking pitch accents (Magne, Astésano, Lacheret-Dujour, Morel, Alter, & Besson, 2005). A related finding on the interpretation of information status from prosody is that the choice of pitch accent melody affects the interpretation of a word as new vs. given (American English: Watson, Tanenhaus, & Gunlogson, 2008; British English: Chen et al., 2007).

The effect of F0 on the interpretation of information status differs across languages, and depends on the relative contributions of word order and prosody in encoding information status. Thus, in Germanic languages, where word order within the sentence is relatively rigid, prosody plays a key role in communicating focus and information status. In contrast, and as noted above for Spanish, many Romance languages and those from other language families use primarily word order to express the same kinds of meaning (Donati & Nespor, 2003). In a carefully designed comparative study of Dutch and Italian, Swerts et al. (2002) show that while Dutch listeners perceive gradient distinctions in emphasis between words bearing focal pitch accents and unaccented words, Italian listeners do not. Furthermore, Dutch listeners but not Italian listeners are able to identify the preceding discourse context of a given utterance based on its pattern of pitch accents. In some so-called free word order languages, word order interacts with prosody in the encoding of focus and information status (Georgian: Skopeteas & Fanselow, 2010; Russian: Slioussar, 2011; Romani: Arvaniti & Adamou, 2011), in which case listeners may consider word order and prosody together in interpreting the information status of a word presented in its discourse context (Russian: Luchkina & Cole, 2013, 2014).

In this section we have seen evidence from a variety of languages, but especially from English and Dutch, for a complex relationship between prosody and focus/information status. Focus and information status are expressed through patterns of relative prominence among elements in a prosodic phrase, such that a word or constituent with special focus (narrow or contrastive) has the greatest prominence relative to other words and constituents in the same phrase. For languages like English where focus is marked through prominence and an associated pitch accent, focused words are phonetically distinguished by the F0 contour that expresses the focal pitch accent, and by duration and intensity as correlates of phrasal prominence. Yet, the interpretation of these phonetic markers as cues to focus or information status is complicated by the possibility that not all pitch accents mark focus or new information. Corpus studies suggest that such mismatches are especially prevalent in spontaneous speech. Other factors related to the phonological, lexical and syntactic context, at a minimum, must be considered in determining focus or information status on the basis of acoustic cues to prominence and pitch accent.

4.2. Prosodic encoding of illocutionary force

At the phonological level, pragmatic meaning related to the illocutionary force of an utterance (e.g., marking an utterance as a statement, question, acknowledgement, etc.) can be associated with individual intonational features and their combination, which determine complex F0 contours over an utterance (English: Beckman & Pierrehumbert, 1986; Gussenhoven, 1984; Pierrehumbert & Hirschberg, 1990; Dutch: Grabe, Gussenhoven, Haan, Marsi, & Post, 1997; Gussenhoven, 1984). For example, a very common intonational pattern for declarative sentences in American English is illustrated in (6a), and its intonational composition is described by Pierrehumbert and Hirschberg (1990) as follows. The sequence of H* pitch accents mark new information on content words, the L-phrase accent expresses completion (the separation of this phrase from what may follow) and the L% boundary tone expresses that the current phrase is not referentially linked to a following phrase, signalling a felicitous end to a discourse segment. The same syntactic form (declarative) can be used to convey a question with a change in the phrasal tones. As described by Pierrehumbert and Hirschberg, a H- phrase accent conveys that the current phrase is part of a larger interpretative unit that includes following material, and the H% boundary tone signals forward reference (cross-speaker, in this example). (6b) could be uttered as a covert request for information or confirmation, for instance, in approaching a receptionist to check in for an appointment:

(6) a. You deliberately deleted my files.
H* H* H* L- L%
b. My name is Mark Liberman
H* H* H- H%

In addition to the production evidence for the prosodic encoding of illocutionary force, there is evidence from a small number of perception studies showing that listeners can identify pragmatic meaning related to illocutionary force based on the intonational features of an utterance (e.g., Italian: D'Imperio & House, 1997; Swedish: House, 2003; Spanish: Face, 2005).

4.3. Prosodic encoding of affective meaning

Particular combinations of intonational features (pitch accents, phrase tones) can also be associated with types of affective meaning, conveying speaker attitude or emotional state related to the interpretation of the utterance. For example, in American English the so-called 'Rise-Fall-Rise' contour over the nucleus of a prosodic phrase expresses the speaker's uncertainty about the utterance in a given context, and in particular, uncertainty about 'some salient relationship between discourse entities' (Ward & Hirschberg, 1985). This contour is comprised of a rising pitch accent (L*+H), which expresses uncertainty or incredulity, followed by a phrase-final continuation rise (L-H%), which expresses incompleteness. The example below (=Ward and Hirschberg's ex. 6) illustrates the use of this contour, which is annotated with \…/ on the accented word, *Bill.* Here the speaker is expressing uncertainty about Bill's status on the scale of sensibility:

(7) How can anyone with any sense not like San Francisco?
B: \Bill/ doesn't like it.

Ward and Hirschberg (1985) analyse the meaning conveyed by the English Rise-Fall-Rise contour as a type of conventional implicature – an aspect of the utterance meaning that is not truth-conditional, but which constrains the appropriateness of the utterance in the discourse context. Culpeper (2011) also points to implicature in his account of prosody in conveying impoliteness in British English. Culpeper illustrates prosodically conveyed impoliteness with the example in (8), an exchange between two young sisters:

(8) A. Do you know **anything** about yo-yos?
B. That's mean.

In this interaction, there is nothing about the lexical content or syntactic form of A's utterance that signals impoliteness. But as Culpeper describes it, speaker A produces this utterance with nuclear prominence on *anything* followed by a sharply falling intonation, which is markedly different from the rising intonation pattern that is typical for yes/no questions in British English. Culpeper describes this as an unexpected mismatch of prosodic pattern and syntactic form, and claims that unexpected prosody carries conversational implicature, triggering a search for pragmatic meaning related to the speaker's attitude. In the interaction in (8), the result of this search is seen in the response from speaker B, who has clearly interpreted A's utterance as intentionally impolite.[10] More generally, Culpeper claims that the discourse meaning of an utterance as conveyed by prosody must be considered in relation to the context. Culpeper, Bousfield, and Wichmann (2003) describe phonological and phonetic effects of prosody that convey impoliteness. In addition to the choice of pitch accents and phrase tones that signal illocutionary force, speakers may select intonational features that signal finality or completion to block the hearer's further contribution, and speakers may also select loudness or pitch range settings as global parameters over the entire utterance, which they suggest may function to control the auditory space and distance the speaker from the hearer.

In work looking at prosodic cues to other types of speaker affect, such as irony and sarcasm, there are inconsistent findings about the role of prosody. In their study of irony in talk radio speech, Bryant and Fox Tree (2005) find that although American English listeners could distinguish ironic and non-ironic speech in bandpass filtered samples, there was no consistent set of acoustic properties that could be identified as signalling an ironic 'tone of voice'. On the other hand, Cheang and Pell (2008) report prosodic correlates of perceived sarcasm in American English in lower mean F0. Comparing sarcasm with other affects such as sincerity and humour shows differences in the harmonic-to-noise ratio (HNR), F0 standard deviation, speech rate and F0 range. The divergent findings in the studies by Bryant and Fox Tree (2005) and Cheang and Pell (2008) may reflect differences in the speech materials judged by listeners. Bryant and Fox Tree used excerpts from spontaneous speech (radio talk shows), while Cheang and Pell used 'posed speech', elicited through reading. It seems likely that the Cheang and Pell's speakers were more consistently recruiting prosody for their explicit task of producing different affects, and that speakers are less consistent in their use of prosodic cues to communicate affective meaning in ordinary speaking conditions. Further research looking at variation in the expression of affective meaning as a function of speech style is called for.

In addition to the works cited above, which focus on prosodic encoding of affective meaning in speech production, there are studies that test listeners' perception of affective meaning as a function of intonation. The general finding is that listeners do make use of prosodic cues of this sort, though with differences in the strength of the association between prosody and meaning depending on the meaning category or the intonational feature tested (English: Gussenhoven, 1984; Dutch: Grabe et al., 1997).

A related finding from Chen, Gussenhoven, and Rietveld (2004) is that listeners' perception of speaker affect from prosodic cues may vary according to the listener's native language, reflecting the normal phonetic implementation of prosody in their language. In that study, Dutch and British English listeners show different sensitivity to variation in F0 range as a cue to speaker attributes like 'confident', 'friendly' and 'emphatic', which the authors relate to the difference in standard pitch range for the two languages (Dutch having a smaller standard pitch range than British English).

4.4. Prosody as a resource for managing interaction

Prosodic features, and intonation in particular, are used in interactive speech to manage talker turn changes, and more generally, to convey cohesion between an utterance and preceding speech. Wennerstrom (2001), in a review of Brazil (1985) and other work on prosody in conversational interaction, describes 'tone concord', as the pitch range setting a speaker chooses at the start of a turn, to match the pitch at the end of previous speaker's turn. Tone concord is described as expressing a speaker's agreement with the perspective of the interlocutor. The absence of concord is described as conveying reproach, astonishment and other discordant stances. A related finding is based on acoustic evidence of pitch concord in backchannels. Heldner, Edlund, and Hirschberg (2010) study backchannels (*mm-hm, okay, yeah*) in American English dialogue and find that the pitch of backchannels is matched to the pitch of the immediately preceding utterance, marking the backchannel as unobtrusive, which arguably facilitates discourse flow.

Prosody is used to convey a range of communicative intentions in dialogue beyond talker turn changes. Recent studies have investigated the role of prosody associated with affirmative cue words like 'okay' in conveying discourse meaning in varieties of English. Gravano, Hirschberg, and Beňuš (2012) investigate affirmative cue words in a corpus of task-oriented dialogue in American English. These words have a variety of discourse and pragmatic functions, such as marking a new topic, marking the beginning and ending of a discourse segment, expressing agreement with the interlocutor and as backchannel speech that acknowledges the continuation of the interlocutor's talk. Gravano and colleagues construct statistical models of the prosodic correlates of discourse/pragmatic function for the affirmative cue words in their sample, and find that word-final intonation (pitch patterns) and word-level intensity distinguish among the various functions of the affirmative cue words. They also report results from a machine learning experiment that uses acoustic prosodic measures, in combination with lexical and discourse features and other phonetic features, to identify discourse/pragmatic function. They show that acoustic prosodic measures make a small contribution to correct classification.

van Zyl and Hanekom (2012) provide evidence that listeners are able to identify communicative intention from prosodic cues associated with affirmative cue words in English. They collected productions of 'okay' under two discourse conditions, expressing the speaker's compliance vs. reluctance to an interlocutor's proposal in the preceding discourse.[11] Listeners subsequently identified these single-word productions presented in isolation as conveying the speaker's compliance or reluctance. van Zyl and Hanekom examine acoustic measures of prosody based on F0, intensity, voice quality and duration, and find that increased duration is the strongest predictor of perceived speaker reluctance. Other acoustic measures of prosody show substantial variation among speakers in their prediction of speaker reluctance.

In addition to its use in conveying a speaker's communicative intentions, prosody can also be affected through entrainment (also *convergence, alignment,* or *imitation*) of one speaker's productions to those of the interlocutor in interactive dialogue (American English: Levitan, Gravano, & Hirschberg, 2011; Levitan & Hirschberg, 2011). Prosodic entrainment may not be intentional on the part of the speaker, and indeed, it is shown to occur in imitation of recorded speech in the absence of any interpersonal interaction (American English: Cole & Shattuck-Hufnagel, 2011). But as shown by Levitan et al. (2012) for American English task-oriented dialogue, prosodic entrainment has social significance and is a predictor of the perceived success of an interaction. In their study, interactions between pairs of speakers that exhibit entrainment are judged by others to have greater efficiency and flow (for male-male interactions) and to be associated with positive social behaviours such as giving encouragement and trying to be liked.

4.5. On the strength of prosodic cues to discourse meaning

Typically, prosody combines with other linguistic information, e.g., lexical and syntactic information, to convey the illocutionary force of an utterance. Sridhar, Bangalore, and Narayanan (2009) show evidence that prosody makes an independent contribution to cueing discourse meaning through an automatic classification study with data from a spontaneous speech corpus. This study uses lexical, syntactic and acoustic-prosodic features to predict manually transcribed dialogue act tags for utterances from the MapTask and Switchboard speech corpora. The results of their classification experiments show that accuracy improves, though by only a small margin of 2.8%, with the addition of the acoustic-prosodic features to the model. The small boost from adding acoustic-prosodic cues suggests that there is substantial redundancy between the dialogue act information conveyed through prosody and information that comes from knowing the lexical and syntactic properties of the utterance. The authors show that further improvement comes with modelling the dialogue act of the preceding context. This study highlights an important point about the function of prosody in communication – when considered over a large body of speech, prosody is only one of the many factors that combine to convey pragmatic meaning. Across speakers, utterances and situations there is substantial variability in the specification of prosodic phonological features to encode discourse meaning (similarly, for encoding structural information at the syntactic or discourse level), and there is further variation in the phonetic implementation of those prosodic features. All of this variability means that prosody is not a fully reliable cue to discourse meaning – though certainly, when prosodic cues are present, listeners attend to them and comprehend meaning according to the information that prosody signals. We return to consider prosodic variability in Section 7.2.

4.6. A biological basis for the prosodic encoding of discourse and affective meaning

Gussenhoven (2002), building on work by Ohala (1983, 1984), argues that variation in the phonetic implementation of intonational features, e.g., in pitch range, the scaling of pitch accents and pitch reset at phrase onset, has its origins in biological factors that relate speaker attributes of size and effort to their effects on vocalisation, and on pitch in particular. For example, a speaker who is small is less likely to pose a threat than a speaker who is large, and since small speakers typically have higher-pitch voices, the association between the speaker attribute (non-threatening) and high pitch is generalised in language, resulting in patterns like the stylized use of high pitch to convey friendliness, politeness and vulnerability, as opposed to the opposite traits, which are likely to be communicated by low pitch. Another example is that of a speaker who exerts great effort while speaking, reflecting a high level of emotional involvement, which results in speech produced with a larger F0 range. Languages generalise this pattern in the association of expanded pitch range with grammatical focus, calling the listener's attention to a word that the speaker marks with greater effort, translated into expanded F0 range.

The claim that there is a biological basis underlying the use of prosodic features to convey linguistic meaning rests on the understanding that the same phonetic parameters that express prosodic features also play a role in the expression of the speaker's physiological and emotional condition. We consider first the effects of speaker's physiological state on the acoustic properties of their speech. Porges (2011) discusses the functions of the parasympathetic nervous system in which the mechanisms that regulate (physiological) stress have effects not only on the heart and lungs, but also on the neuro-muscular systems of the ear, larynx and pharynx. States of calmness and safety are associated with increased neural tone to the laryngeal and pharyngeal muscles resulting in speech with lower frequency and increased frequency modulation at lower frequencies, while states of danger and distress are associated with higher pitched vocalisations. Porges et al. (2013) argue that this model can explain, among other things, the dampened vocal prosody and other manifestations of impaired social engagement in children diagnosed with autism spectrum disorders. Other studies have looked for evidence that speaker emotion is reflected in prosody. For example, Pell, Paulmann, Dara, Alasseri, and Kotz (2009) show that F0 mean, F0 range and speech rate are significantly different for seven perceptual categories of emotion (as identified by native listeners) for English, German, Hindi and Arabic. These prosodic cues to emotion were sufficient to allow native listeners to perceive the emotion (as identified from elicitation criteria) in de-lexicalized speech with accuracy ranging between 59–81%, which was well above the chance level of 14% in this study. A similar finding for F0 correlates of emotional arousal is found in the study of German by Ladd, Silverman, Tolkmitt, Bergmann, and Scherer (1985).

To summarise here, prosody is influenced by speaker's affect or emotional state in patterns that may relate to a more fundamental biological 'code'. Further research is required to understand how, from a common biological foundation, languages may come to differ in the association of speaker affect with prosody, and why listeners exhibit only moderate accuracy in perceiving speaker affect or emotion based on prosody and other vocal cues.

5. Prosodic relativity

It has been noted several times in the preceding sections that the acoustic cues to prosodic prominence and boundaries are interpreted *relative to* the prosodic features of neighbouring words, syntactic boundaries or discourse units. In light of the possibility of such dependencies between prosodic features, it is appropriate to say that in some cases, prosody signals relationships between words, syntactic phrases and discourse units that reveal properties of the broader grammatical and discourse context. This section reviews several examples of dependencies among prosodic features that constrain their phonetic realisation and/or interpretation.

5.1. Context effects on the alignment of prominence-lending F0 contours

To illustrate the role of context in constraining the phonetic realisation of prosodic features, we need look no farther than the phonological context of pitch accent. In many languages the F0 realisation of a pitch accent can be limited by the presence of another intonational tone on the accented syllable, or the syllable following it. For example, in American English, when the nuclear pitch accent in an utterance occurs on or near the final syllable in the prosodic phrase, the F0 peak of a nuclear pitch accent is typically retracted (i.e., is located earlier in the accented syllable), in comparison to the F0 peak of a pre-nuclear pitch accent that is positioned earlier in the phrase. Silverman and Pierrehumbert (1990) analyse this pattern of early alignment of the accentual peak as an effect of tonal crowding in the nuclear accent condition. The upcoming phrase accent (an edge-marking tone) encroaches on the nuclear pitch accent, and retraction of the accentual peak is one way of increasing the distance between the accentual F0 peak and the F0 contour defined by the edge-marking tone. A similar pattern of accentual F0 peak retraction before a prosodic boundary is reported by Grabe, Post, Nolan, and Farrar (2000) for Southern British English, and by Prieto, van Santen, and Hirschberg (1995) for Mexican Spanish. Arvaniti, Ladd, and Mennen (2006) report a parallel retraction of the low F0 turning point of a nuclear L* accent in Greek. Not all languages resolve tonal crowding by retraction of an accentual F0 peak or valley. In similar conditions, Leeds British English (Grabe et al., 2000), Hungarian (Ladd, 2008, p. 182) and Palermo Italian (Grice, 1995) exhibit truncation of an F0 contour, with loss of the final portion.

Context at the level of discourse structure can also affect F0 alignment. Wichmann, House, and Rietveld (2000) compare F0 peak alignment for pitch accents in relation to the position of the accented word in the topic units or paragraph, and find that while pitch accents display *early* F0 peak alignment in the *final* position of a discourse unit (as predicted from studies showing early alignment for phrase-final nuclear accents), there is an opposite pattern of *late* F0 peak alignment for pitch accents that are *initial* in the discourse unit. It is not clear if the late peak alignment in discourse-initial position reflects tonal crowding, e.g., from a phrase- or topic-initial tone (a 'paratone'), but this pattern does present another example of a contextual effect on the phonetic realisation of pitch accents.

What the patterns described above show is variation in the segmental alignment of F0 peaks, or variation in the completion of an F0 contour, depending on the preceding or following context. In order for any of these variant F0 contours to be recognised as cues to underlying pitch accents and boundary tones, the F0 patterns must be interpreted relative to the local context, at the level of the prosodic phrase, and at higher levels of discourse structure. A phonological pitch accent may be realised with an early F0 peak in phrase-final position, with a later F0 peak in phrase-medial position, and possibly with even later peaks occurring initially in a topic unit or paragraph. Therefore, peak alignment as a cue to pitch accent type must be considered in relation to the position of the accent in relation to its position in the prosodic phrase and in the discourse unit.

5.2. Relative peak height in downstep

Another example where the local grammatical context must be considered in the interpretation of a prosodic feature is downstep – the step-wise lowering of an F0 peak associated with a High tone, such as a H* pitch accent or a H- phrase accent in the ToBI system.[12] Typically, downstepped tones are preceded by a High tone, and it is in reference to the preceding F0 peak that the downstepped tone is identified (Ladd, 2008, pp. 97–99, 105–106, *passim;* Yoon, 2007). In English, downstepping can occur across any sequence of high-tone pitch accents (e.g., H*, L+H*), or across a sequence of a high-tone pitch accent followed by a H- phrase accent, and when present conveys 'a nuance of finality or completeness' in English (Ladd, 2008, p. 78). A similar finding is found for Dutch (Swerts et al., 1994).[13] A downstepping pattern is typical across the H* pitch accents in word sequences that comprise a list (e.g., 'black, blue, red, green, and white'), resulting in a downward-trending, terraced F0 contour. Downstep is also common in English in the stylized 'calling contour', as in (9), where the exclamation point marks the downstepped tone (!H*):

(9) H*!H*
Taxi's waiting!

Downstep is typically bounded by the prosodic phrase, and so downstepping patterns provide a cue to prosodic phrase structure. Evidence from perception studies confirms that listeners use downstep (and pitch reset at the end of a downstep sequence) as a cue to prosodic phrase boundary (de Pijper & Sanderman, 1994; Sanderman & Collier, 1997).

Truckenbrodt (2004) offers further evidence of local dependencies between pitch accents in downstep. He argues on the basis of F0 patterns in German that there are two components to downstep, each of which illustrates an effect from prosodic context: (1) in a sequence of two L*+H (rising) pitch accents, the F0 peak of the second pitch accent is lowered relative to the preceding peak and (2) the F0 peak of the L*+H pitch accent is *upstepped* (^H) in the context of a following downstepped High tone. These two effects combine in a sequence of two rising pitch accents, L*+^H L*+!H, to derive a pattern of downstep enhancement, where the raising of the first peak serves to enhance the distinctive lowering of the downstepped peak that follows.

The German and English downstep patterns illustrate what Ladd (2008, pp. 304–308) refers to as ‘a syntagmatic relation of pitch level between two accents or other prosodic constituents’, which he analyses in terms of a metrical structure that assigns a high–low or low–high pattern of pitch register features to the two prosodic elements in the metrical structure. Successive pitch accents parsed in a high–low (register) metrical structure will exhibit downstep of the high tone in the second accent.

5.3. Relative boundary strength

Syntagmatic relations also come into play in determining the relative strength or degree of prosodic phrase boundaries and prominence. As discussed in Section 3.2, prosodic phrase boundaries tend to occur at major syntactic boundaries, and therefore, the presence of a prosodic boundary offers a cue to the local syntactic context. Moreover, several studies have shown that the relative strength of prosodic boundaries internal to a complex sentence can disambiguate between competing syntactic structures for the sentence (Wagner, 2010; see Wagner & Watson, 2010 for a review). For example, Ladd (1988) finds that in syntactic structures with coordination and disjunction such as *X and Y but Z*, speakers of British English produce internal prosodic boundaries after both the *X* and *Y* phrases, but with a stronger prosodic boundary at the end of the phrase that is higher in the syntactic tree, in this example after the *Y* phrase. So, the syntactic structure [[*X and Y*] but Z] is produced as *X* | *and Y* || *but Z* (where | designates a weaker prosodic boundary and || designates a stronger one). Ladd (2008, pp. 293–297) argues that the difference between the weaker and stronger boundaries is not a categorical difference in boundary type – there is no absolute property that distinguishes boundaries in one syntactic context from those in another – but rather that prosodic phrases can be recursively nested and boundary strength decreases with depth of embedding. What is significant for the present discussion is that it is the *relative* strength of two or more boundaries in a complex sentence that cues the syntactic relations among the phrases, so interpreting the prosodic boundary as a cue to the syntactic context requires consideration of a prosodic boundary in relation to its prosodic context.

Similar evidence for the importance of relative boundary strength is shown by Peppé et al. (2000), who examine the production of prosodic boundaries in Southern British English in sentences with the pattern *X Y and Z,* elicited as descriptions of 2-item lists (*cream-buns and cheese)* or as 3-item lists (*cream, buns, and cheese*). Speakers in that study were fairly consistent in distinguishing these structures on the basis of the relative strength of prosodic boundaries after the *X* and *Y.* Specifically, silent pauses located following the first word (*cream*) and the second word (*buns*) were indicative of a 2-item list [[X Y]] and Z], but only if the second pause was longer than the first. In producing the 3-item lists, speakers were more likely to produce pauses after the first and second words that were equal in duration, or with the first pause having longer duration than the second. In other words, the relative duration of pauses as cues to prosodic boundaries was a direct reflection of the relative embedding of the first and second words in the sequence. The authors of that study state that ‘for all prosodic elements, the relationship between the exponency on the two parts was important’. (Peppé et al., 2000, p. 322).

The relative strength of a prosodic boundary may also signal differences in discourse segmentation. For instance, in their study of spontaneous narratives in Mandarin, Tseng, Su, and Lee (2009) show that boundaries marking the end of discourse units at the level of ‘paragraph’ and within-paragraph ‘breath groups’ are distinguished by the relative value of acoustic measures such as the intensity of the speech immediately preceding the prosodic boundary. The authors claim that the combined set of relative prosodic cues, reflecting global prosodic organisation, is more successful in discriminating discourse boundaries than any single cue considered on its own.

5.4. Relative degree of prominence

There is also evidence for prosodic relativity in the prosodic encoding of focus through prominence assignment. In American English, words with narrow or contrastive focus and discourse-new words may both be marked as prominent and assigned a pitch accent, yet studies show that speakers can distinguish between these two types of prominence on the basis of relative prominence, and in the relative scaling of the F0 contour expressing the pitch accent (Katz & Selkirk, 2011; Xu & Xu, 2005). For example, Katz and Selkirk (2011) look at the prosodic distinction between contrastive focus and discourse new constituents in American English, in sentences like ‘*But they only speak* ***Wolof*** *in* ***Mali***’ produced with prominence that marks contrastive focus on one of the bolded words, and prominence that marks discourse-newness on the other. They had speakers produce three productions of each sentence in different discourse contexts that elicited three patterns of focus/new marking on the first and second complement of the verb: Focus-New, New-Focus and New-New. Evidence from F0 and duration measures show that the two types of prominence differ in the relative size of the pitch rise and fall of the first complement (*Wolof*) compared to the same measures from the second complement (*Mali*). Katz and Selkirk argue that the difference cannot be attributed to a categorical difference in pitch accent type (e.g., H* vs. L+H*), nor does it depend on a difference in prosodic phrasing. Rather, interpreting a complement as bearing contrastive focus, or as discourse-new requires evaluating the acoustic cues to prominence on that word *relative to* the same cues on the other complement.

Prosodic relativity is also observed in the scaling of prominence-lending pitch accents depending on the position of the accented word in the discourse. For example, Swerts and Geluykens (1993, 1994) find higher F0 peaks for prominence-lending pitch accents on words that introduce new topics, and a pattern of declining F0 across the utterances within a topic unit. Swerts and Geluykens (1994, p. 31) remark that the variation in F0 peak height 'points to a very sophisticated use of global F0 features by the speaker, and shows that we should also look beyond the local level when studying the discourse functions of F0 variation'. Similarly, Menn and Boyce (1982), in their study of American English dialogues between adult and child speakers, find a higher peak F0 value in utterances that mark the start of a new topic.

Just as Katz and Selkirk find production evidence that a difference in relative scaling of pitch accents distinguishes contrastive focus and discourse newness, Krahmer and Swerts (2001) find similar evidence of prosodic relativity from their perception study with Dutch listeners. Krahmer and Swerts find that Dutch listeners perceive a difference in degree of prominence between a word with a pitch accent marking contrastive focus and the same word with a pitch accent marking its status as discourse new in a matched utterance (e.g., comparing the adjective with contrastive focus in 'the **BLUE** SQUARE' with the same adjective with new-information focus in 'the BLUE SQUARE'). Of interest, the perceived difference in prominence between these two accent types disappears when the words are taken out of their sentence context and presented in isolation. In that condition, listeners did not reliably perceive a distinction between pitch accents marking contrastive focus and those marking discourse-new status.

There is also evidence for prosodic relativity in the perception of F0 patterns of prominence-lending pitch accents, where peak height and relative prominence is judged in relation to the position of the accented word in the utterance. As noted in Section 4.1, Dutch listeners judge the relative height of two F0 peaks in the same utterance in distinguishing between single- or double-focus interpretations (Rump & Collier, 1996).The same perceptual effect is found in studies that ask listeners to judge prominence without an explicit reference to information status: listeners judge F0 peaks that occur later in an utterance as more prominent that peaks of equal height that are located earlier in the same utterance (English: Pierrehumbert, 1979; Dutch: Gussenhoven & Rietveld, 1988). This effect of pitch accent position is also observed in Dutch utterances that contain only a single pitch accent (Gussenhoven, Repp, Rietveld, Rump, & Terken, 1997). A striking finding from American English is that listeners' sensitivity to the difference in peak height between a pair of successive pitch accents is different when F0 peak height increases across the two pitch accents in an upward trend than when F0 peak height decreases across the pair (Ladd & Morton, 1997). While the basis for the order effect is not fully understood, on the whole these findings suggest that listeners compensate for a declining F0 baseline across the utterance in interpreting degree of prominence.

6. Prosody signals information from the situational context

All of the dependencies relating prosody and context mentioned so far result from grammatical conventions that associate a particular prosodic form with a property of the lexico-syntactic or discourse context. This section discusses several ways in which prosody reveals extra-grammatical properties of what we may call the situational context, including speaker attributes, as well as the speech style as related to the communication setting.

6.1. Prosodic indexing of speaker identity

One of the surprising findings to emerge from prosody research is the high degree of individual speaker differences in the production of prosody. Few studies to date have examined individual differences as a primary focus of investigation, but several studies report such differences, and other studies reveal individual differences in descriptive statistics, without commenting on the finding. Individual differences are surprising precisely because of the heavy load prosody carries in signalling meaning at so many levels of linguistic specification, as reviewed in the preceding sections. Given that prosody has the capacity to convey information about the grammatical and discourse context that is critical to the intended meaning of the utterance, and given that listeners are sensitive to prosodic cues in comprehending speech, it is remarkable that speakers are not more consistent in the expression of prosody. Of course, the fact that prosody performs many functions may itself be a reason for individual differences – depending on which aspect of meaning is deemed more important for the communication goal, a speaker may choose to commit phonological and phonetic resources to prosodic encoding of one linguistic function at the expense of another. Following is a sampling of individual speaker differences as reported in the prosody studies already cited in this article, which reveals speaker-dependent variation across a wide range of prosodic functions.

6.1.1. Focus

Peppé et al. (2000) report variation among British English speakers' production of corrective focus elicited by the experimenter through prompting questions, with variation in the type of pitch accent produced on the focused word, and in the phonetic details related to silent pause, loudness and lengthening. de Jong and Zawaydeh (2002) find differences among four speakers of Ammani-Jordanian Arabic in their phonetic implementation of focus in the production of read sentences. This contrasts with the relatively greater consistency among the same speakers in the pattern of acoustic correlates of word-level stress. The authors attribute the lesser consistency in the phonetic expression of focal accent compared to stress to greater degree of conventionalization for stress. Focus, unlike word-stress, is more likely to be influenced by the broader discourse context, which could lead to greater variability across speakers.

6.1.2. Illocutionary force/speech act

Grabe (2004) finds speaker variability in the choice of pitch accent melody that conveys four discourse functions (declarative, Wh-question, yes/no-question, declarative question) across six speakers for each of seven dialects of British English performing a speaking task involving read sentences.

6.1.3. Discourse and rhetorical structure

Swerts and Geluykens (1994) find variability among three Dutch speakers in their production of prosodic cues to discourse structure in interactive, instruction-giving dialogues. Speakers differed in their use of a low F0 to mark end of topic, in marking new topics with the highest F0 peak and in the pattern of declining F0 over topic units. Of interest, speakers were consistent in their use of pause at locations of topic shift, and in the relative duration of those pauses with pauses internal to topic units. Den Ouden et al. (2009) investigate prosodic marking of rhetorical structure in read aloud news reports, and find variability among 20 Dutch speakers in their use of pause and F0 to mark hierarchical relations between different segments of the text. The authors speculate that these differences may correspond to differences in reading fluency among participants, and/or differences in the choice of F0 and pause as acoustic cues to express rhetorical structure.

Speaker variability in prosody production does not go unnoticed by listeners. Listeners are able to use information from prosody as an aid to comprehension, but only to the extent that speakers produce salient prosodic cues. Consider again the study by Swerts and Geluykens (1994) examining prosodic marking of discourse structure in Dutch. Swerts and Geluykens tested listeners' ability to detect end-of-topic locations in the speech from three speakers (as described above), applying band-pass filtering to render the speech samples unintelligible and manipulating F0 and pause duration to selectively neutralise those cues in some samples, while retaining the speaker's original F0 and pause cues in other samples. Listeners' responses show sensitivity to prosodic cues to discourse structure, and notably, response accuracy varies by speaker in ways predicted by the differences among speakers in their use of F0 and pause to mark discourse structure. This finding underscores the role of prosody in marking discourse structure, and the variable strength of prosodic cues due to speaker variability.

Considering that listeners attend to prosodic cues to grammatical and discourse context, and given that speakers vary in their production of those cues, it follows that listeners may be sensitive to speaker-dependent patterns of variation in prosody. Evidence for this comes from studies of variation in listeners' perception of prosodic prominence and boundaries in spontaneous, conversational speech. In a large-scale study of prosody perception, the present author and her colleagues asked untrained listeners to identify prosodic boundaries (described in terms of 'chunking' that helps the listener interpret the utterance), and prominent words (described as 'highlighted for the listener' and 'standing out from other non-prominent words') in speech excerpts from 36 speakers from the Buckeye corpus of American English spontaneous speech (Pitt et al., 2007). The findings show that listeners' perception of prominence and boundary is correlated with acoustic measures of duration, intensity, and to a lesser degree, F0 (Cole, Mo, & Baek, 2010; Cole, Mo, & Hasegawa-Johnson, 2010). Of interest here is the finding that the strength of these acoustic cues to prominence and boundary substantially increases when the acoustic measures are normalised within-speaker and within-discourse segment (Mo, 2011). This strongly suggests that listeners are taking into account the individual speaker's prosodic habits in that discourse segment and judging the prosodic features of a given word against the backdrop of those habits.

Listener sensitivity to the individual speaker's pattern of prosody production suggests that a listener who is sufficiently familiar with a speaker may be able to use information from prosody to identify a speaker. This possibility is further suggested by work on automatic speaker identification (also called speaker verification in the literature), using acoustic prosodic measures from primarily F0, but also including intensity to identify the speaker in excerpts from the Switchboard corpus of telephone conversation speech (Adami, Mihaescu, Reynolds, & Godfrey, 2003; Farahani, Georgiou, & Narayanan, 2004; Weber, Manganaro, Peskin, & Shriberg, 2002). Each of the cited studies reports lower error rates in speaker identification when acoustic correlates of prosodic features are used. These results alone do not tell us if human listeners rely on acoustic prosodic measures to identify a speaker, for example, in situations like telephone conversation which lack visual information, but they suggest the prosodic cues have the potential to aid in speaker identification, even when considered by themselves.

Related to the question of how speaker identity is revealed through acoustic prosodic cues is the question of how prosody encodes speaker attributes such as sex or gender. The most obvious effect of speaker sex on prosody is in F0, with males typically producing lower F0 than females. Beyond this obvious difference, studies of New Zealand English (Daly & Warren, 2001; Warren & Daly, 2000) and Dutch (Haan & van Heuven, 1999) find other differences between male and female speakers. Females exhibit larger dynamic F0 range (both studies) and larger F0 rise excursions (New Zealand English) than male speakers. Female speakers also have a greater tendency to use marked or mismatched prosody, such as the high-rising contour at the end of a declarative (Warren & Daly, 2000). These sex- or gender-based prosodic patterns can serve as cues to the sex and gender of the speaker as a property of global context, in situations where visual cues are indeterminate or absent.

6.2. Prosodic indexing of speech style

There is some evidence to suggest that speakers produce different prosodic patterns when reading as compared to spontaneous speech. Blaauw's (1994) study, mentioned in Section 3.2, compares the prosodic features of spontaneously produced instruction monologues with read-aloud productions of the transcribed monologues by the same speaker, and finds differences in the distribution of prosodic boundaries and their acoustic realisation. Hazan and Baker (2010) observe higher median F0 in English read speech utterances compared to interactive speech produced in clear or casual styles, while Swerts, Strangert, and Heldner (1996) report steeper declination and stronger F0 resets in Swedish read speech. Of course, professionals trained in public speaking or news announcers may produce very fluent and natural prosody even when reading, but impressionistically is seems that many ordinary speakers produce different prosodic patterns when reading than when speaking spontaneously (cf., Hirschberg, 2000). From my reading of the literature, it is not evident that anyone has tested if listeners can detect speech style based on prosodic cues, but if speech style is sufficiently and consistently distinguished in the production of prosody, it can be expected that listeners will make use of that information whenever it is available. This reasoning leads us to predict that prosody may provide cues to speech style as a property of the situational context.

7. What we have learned, grand challenges and the way forward

The research reviewed above demonstrates that prosody and context are interrelated at every level of linguistic organisation from the word up to the discourse segment. One aspect of this dependency is in the prosodic encoding of properties of the linguistic context at the level of the word, the (syntactic) phrase, the utterance and the discourse or rhetorical unit. Sometimes prosody provides a redundant cue to information that is also conveyed through words or morpho-syntactic structure, and sometimes prosody is the sole expression of a grammatical or discourse feature. The second aspect of the dependency between prosody and linguistic context is in the effect of context on prosodic form and on the interpretation of prosodic features. Looking beyond grammatical and discourse context, prosody also conveys information about the broader context of the communicative situation, indexing the speaker attributes and speech style and possibly other information. The studies cited here, and many others, show that prosody conveys contextual information through a number of distinct acoustic parameters – including F0, intensity and tempo, and also through acoustic parameters that realise segmental contrasts, which may be hyperarticulated in positions of prominence, and hypoarticulated in prosodically weak positions. There is also compelling evidence for a phonological representation of prosody that mediates between, e.g., syntactic boundaries and their prosodic marking through pause or final lengthening, or between focus and its expression in terms of F0 or duration. The emerging view is that of layered prosodic phrase structure and metrical structures through which prominence relations are defined.[14]

Earlier in this review, at the end of Section 2, two general questions were posed about the relationship between prosody and context, repeated here:

- What aspects of prosody are affected by context?
- What kinds of contextual information are conveyed through prosody?

From the research reviewed above, we have seen that context affects the *location, degree and tonal melody* of prosodic prominence at the level of the word, phrase, utterance and discourse unit. Similarly, context affects the same properties of prosodic boundaries at the level of the phrase, utterance and discourse unit. The specific findings reviewed here are summarised as follows. Context related to the *lexical, syntactic and discourse structure* influences the location of the prominent element within a prosodic constituent and the boundaries of those constituents. Syntactic and discourse structure also determine the strength of a prosodic boundary relative to other nearby boundaries. On the other hand, *focus or information status*, properties that relate to discourse meaning, affect prosody at the level of the phrase and above, in the location of prominence and prosodic boundaries and in the degree of prominence relative to other prominent elements in the same prosodic domain. *Illocutionary force, affective meaning* and the speaker's *communicative intention*, which are additional components of discourse meaning, affect the tonal melody associated with prominent elements and prosodic boundaries at the phrase level, with an added effect of communicative intention on acoustic intensity. Properties of the *situational context* affect *global aspects of prosody*. Specifically, speech style (read vs. spontaneous), speaker sex, speaker emotion and interlocutor interaction impact F0 range, with speech style showing additional effects on F0 declination. In addition, the speaker's emotional state affects F0 register, range and speech rate.

The number and variety of associations between prosody and context, and the many distinct acoustic and articulatory correlates of prosody combine to make a dizzying, diverse range of prosodic effects of context. Sections 7.1 and 7.2 discuss variability, data sourcing and transcription as factors that pose persistent challenges for researchers working with prosodic data, Section 7.3 discusses an approach to the analysis of variability, and Section 7.4 offers some desiderata for future research.

7.1. Prosodic variability

Adding to the complexity is the observation that prosody is variable – prosodic encoding of at least some grammatical factors (e.g., syntactic phrase boundaries) is optional, and the phonological features and phonetic expression of prosody may also vary across speakers. As Hirschberg (2002, p. 32) states:

> research on prosody, as on many linguistic phenomena which rely upon context for their interpretation is more a matter of finding likelihoods – not simple mappings from syntax or semantics or even from an underlying meaning representation to a clear set of prosodic features, for any sentence.

The variability of prosodic encoding can make it very difficult to identify the prosodic features of an utterance or to establish the validity of a phonological prosodic transcription, especially for spontaneous, casual speech, which in turn makes it difficult to chart the relationship between the prosodic features of an utterance and its grammatical and situational context.

One factor that may contribute to variability in the phonetic expression of prosody is the interaction among prosodic effects due to different contextual factors. For example, it is known for English that the F0 pattern of a word varies according to its focus, information status, position in the prosodic phrase, position in discourse structure, topic status or due to speaker attributes such as sex and gender. What is not so clearly understood is how those factors combine to influence the F0 on a specific word within a specific utterance. There are a handful of production studies that have explicitly investigated the interaction among factors in their effects on acoustic correlates of prosody, e.g., examining the interaction between *word-level stress* and *prosodic phrase boundary* in the kinematics of lip and jaw movements in English (Edwards et al., 1991), in F0, duration and formant measures in Jordanian Arabic (de Jong & Zawaydeh, 1999), and in acoustic correlates of Dutch stop consonants (Cho & McQueen, 2005). Interaction between *word-level stress* and *phrasal prominence* (focus) is also observed in Jordanian Arabic (de Jong & Zawaydeh, 2002), and in Cho and McQueen's Dutch study. The same Dutch study shows an interaction between *prosodic boundary* and *phrasal accent*, which is further supported with evidence from syllable duration in Dutch study (Cambier-Langeveld, 1999). Two of the above studies also report a three-way interaction among stress, phrasal accent and boundary (Cho & McQueen, 2005; Edwards et al., 1991).

7.2. Sources of data and the problem of transcription

These findings of interactions among factors that influence prosody are important, though they cover only three of the many grammatical factors that influence prosody. Further research is needed to expand our understanding of the interactions among the full set of grammatical factors, and bringing in discourse and situational factors. Work along these lines will require experiments with complex design, and given the inherent variability in prosodic encoding, will likely require large amounts of data to ensure results with sufficient statistical power. More critically, to discover interactions involving discourse or situational factors will require expanding research methods beyond read speech, to include elicitation of interactive speech in contexts where speakers are engaged in genuine communication with meaningful discourse goals. Examples of prosody research based on interactive speech materials are already in place, including for example, Swerts and Hirschberg (2008; and numerous other works by these authors individually and together), Edlund and Heldner (2005) and others. The call for research on prosody in interactive speech is not to deny the value of experiments with read speech, which allow the experimenter to exert greater control over speech materials towards the desirable goal of reducing variation in speech across test utterances (Xu, 2010). But interactive, goal-oriented speech is necessary to establish discourse and situational context, and may also promote speakers' production of prosodic cues to local grammatical structure. Moreover, some prosodic functions are specific to interactive dialogue, such management of turn-taking or floor-holding strategies (Heldner & Edlund, 2010; Local, Kelly, & Wells, 1986; Selting, 2000), so interactions involving these prosodic patterns must be investigated with interactive speech materials. Possible differences between the prosody of read vs. spontaneous speech, noted in Section 6.3, further suggest the importance of investigating interaction among prosodic effects across a range of speech styles.

A serious challenge confronting prosody researchers interested in the relationship between prosody and context is the need for data that is annotated for prosodic features and for the syntactic, discourse or other contextual features that are thought to relate to prosody. Prosodic annotations are typically obtained from experts who are trained in the use of a particular transcription method, but transcription is known to be difficult and slow, with time estimates ranging from 10 to 100 times the duration of the audio speech recording (Cole & Hasegawa-Johnson, 2012). Transcription is based on the subjective auditory impression of the transcriber, augmented with the visual display of the speech waveform, spectrogram and pitch track. To establish the validity of the transcription two or more transcribers must independently transcribe the same materials sampled from the database, and the agreement rate between the two defines the upper bound for accuracy in the subsequent analysis of prosody in the speech materials. Studies that report inter-transcriber agreement for prosodic transcription demonstrate that while agreement on the presence/absence of a prosodic feature can be quite high, agreement on the degree or type of feature (boundary or prominence) is quite a bit more variable (English: Pitrelli et al., 1994; German: Grice, Reyelt, Benzmuller, Mayer, & Batliner, 1996). It is notable that statistics are not routinely reported in prosody research that relies on transcribed data, raising questions about validity of the proposed analyses. Prosodic transcription is more difficult with spontaneous speech than with read speech, due to the greater variability and the frequent occurrence of disfluency with the former. Researchers who want to collect new speech data to investigate context effects on prosody must have the resources to invest in creating prosodic transcriptions for their data, which can be a substantial burden.

To relate prosody to syntactic and discourse context requires analysing the co-occurrence patterns for prosodic and syntactic, or prosodic and discourse features. The syntactic features of an utterance should be in principle less difficult to label than the prosodic features, but in practice even syntactic annotation can be problematic for spontaneous speech, due to run-on sentences and sentence fragments that make transcribing anything more than a shallow parse structure a challenging undertaking. Discourse annotation is even more difficult (Hirschberg, 2002). Like prosody annotation, it depends in part on the subjective evaluation of the transcriber who must determine the structure of the discourse, e.g., identifying topics, the boundaries of discourse segments, the function of an utterance in the given context and the speaker's attitudes and communicative intentions, among other contextual features. These features can often be inferred from the design of a controlled experiment, but are challenging to identify and validate in spontaneous speech. For example, in a large corpus study of spontaneous Swedish dialogues, Edlund, House, and Strömbergsson (2012) carried out an analysis of transcriber annotations of question types (Yes/No, Wh-, and Alternative), and find that trained transcribers generally agree at rates of 80% and above, with lower agreement for certain question types, including those identified as having a restricted set of Alternative responses. The overall point here is that even trained transcribers assessing discourse function without the pressures of real-time auditory speech comprehension have difficulty in determining discourse function for 20% or more of the utterances in spontaneous, interactive speech. It appears then that further work is needed on discourse annotation for spontaneous, interactive speech before we can expect substantial advances in understanding the role of prosody in communicating the discourse function of sentences and discourse relations among sentences.

7.3. Coping with variation in prosody perception: parsing models

Variability in the phonological encoding of prosody and its phonetic expression raise an obvious question: How do listeners cope with this variability in interpreting acoustic prosodic cues in relation to context? This question is, of course, not unique to prosody. Acoustic cues to phone identity are similarly influenced by many of the same factors from the grammatical, discourse and situational context (e.g., Klatt, 1976). One approach to the problem of cue variability invokes a parsing mechanism in speech perception, by which listeners attribute information in the acoustic signal to known sources of variance (Fowler & Smith, 1986). For instance, English listeners attribute nasalization in a vowel to an upcoming (but not yet perceived) nasal consonant (Fowler & Brown, 2000), factoring the acoustic evidence of nasality from the formant structure of the vowel and identifying its source in an upcoming nasal consonant through the process of coarticulation.

Parsing allows the listener to compensate for coarticulation and other effects of context, through a process of attributing partial quantities in observed acoustic measures to sources that are extrinsic to the segment from which the measurements are drawn. Parsing also allows the listener to make predictions about upcoming context, and provides additional evidence for (or against) the identification of preceding material. In essence, parsing allows the listener to exploit information about context, and is an alternative to normalisation processes which essentially discard such information.

Cole, McMurray, Linebaugh, and Munson (2010) develop a statistical model of parsing in an analysis of vowel coarticulation (see also McMurray, Cole, & Munson, 2011; McMurray & Jongman, 2011). The model handles variability in vowel formants due to coarticulation, and shows that parsing minimises the observed variance in F1 and F2 due to coarticulation with upcoming phones, and variance due to speaker, with the result that the parsed acoustic signal yields better discrimination between vowel categories (phonemes). Parsing requires prior knowledge of the mean and variance of acoustic prosodic cues in distinct contexts (e.g., the vowel /a/ produced before a following /t/), but a listener who has gained that knowledge through prior experience can then use the information to identify the contextual feature in the presence of the cue. Parsing not only allows the listener to make a prediction about the context, but it also affords the listener better discrimination of the linguistic feature associated with the cue.

Parsing can be applied to the problem of variable prosodic cues, as follows. A listener processing the duration of a syllable that is final in a word may attribute the observed long duration in part to the prosodic context by positing a prosodic phrase boundary following the word. In doing so, the listener not only is able to make a prediction about the phrasal context, but is also in a better position to detect a durational cue to prominence in the same region, or a durational cue to a lexical vowel length contrast.

Viewed in this light, variability can be harnessed as a source of information about context that facilitates the perception of acoustic cues to prosody, which should in turn facilitate understanding of the meaning conveyed by prosody. To further develop a parsing model of prosody perception in the face of variable prosodic encoding will require new research on prosody production to establish the mean and variance of individual acoustic cues in known contexts. Also needed are experiments designed to allow the interaction of multiple contextual factors that contribute to the variation of individual cues. Statistical models of the interaction among contextual factors can be constructed to determine the *potential* of an acoustic parameter to signal each individual contextual factor, and must then be tested against behavioural evidence from human listeners.

7.4. Future research

Prosody research over the past five decades (and more) has contributed a wealth of detailed empirical findings on the phonetic expression of prosody, a framework for the phonological representation of prosody and a broad sense of the function of prosody in conveying information related to grammatical and (to a lesser extent) extra-grammatical context. Looking forward, to gain a fuller understanding of the dependencies that link prosody to context, including local and global context, and context related to grammatical features, discourse features and the communication situation, new research is called for that will (1) expand the scope of inquiry to cover a broader range of languages that differ in the prosodic, lexical and syntactic mechanisms used to convey discourse meaning; (2) determine the interaction among factors that condition variation in the encoding of prosody, at the phonological and phonetic levels; (3) reveal individual differences in the production and perception of prosody; and (4) allow the comparison of prosody as produced in different situations, including reading and spontaneous talk, and under different conditions of interlocutor interaction. Some excellent examples of such research are already in place, including works by authors cited here, and especially promising are new efforts that bring together researchers with expertise in different areas that combine to cover the broad field of linguistic context.

Acknowledgements

I thank José I. Hualde and Stefanie Shattuck-Hufnagel for critical input on many of the topics discussed here.

Funding

This work was supported by the National Science Foundation BCS 12-51343, IIS 04-14117 and IIS 07-03624.

Notes

1. The view of prosody as the organisational structure of speech goes back to early works such as Trubetzkoy (1958), as noted by Beckman (1996), and is well established in contemporary linguistic theory, forming the basis for metrical stress theory (Liberman, 1975; Liberman & Prince, 1977; Hayes, 1995) and the autosegmental-metrical theory of intonation (Beckman & Pierrehumbert, 1986; Gussenhoven, 2004, p. 58; Ladd, 2008, ch.8; Pierrehumbert, 1980). An opposing view in contemporary work holds that prosodic features are directly determined on the basis of syntactic or semantic properties of an utterance, denying a role for phonological prosodic structure (Xu & Xu, 2005; see also the discussion in Wagner and Watson (2010, p. 936).
2. Examples of meaning conveyed by (paradigmatically contrastive) prosodic features are plentiful. Prosodic contrast involving tone features at the word level is found in many Asian and African languages. For example, in many Bantu languages the contrast between a High tone and a Low tone on the initial syllable of a verb stem (treated as a prosodic domain) can mark a lexical or grammatical distinction (Downing, 2011). In varieties of English, a similar contrast between High and Low tones at the end of a prosodic phrase (again, a prosodic structure) can signal a difference in pragmatic meaning. As described by Gussenhoven (2004, pp. 296–301), a pitch fall on the final syllable in a declarative sentence signals new information that the speaker is introducing to the discourse, while the same sentence with a slight pitch rise at the end signals that the information in the utterance is already shared by the speaker and listener.
3. Works on English include Nespor and Vogel (1986/2007); Selkirk (1984, 1986); Frazier et al. (2004); Watson and Gibson (2004); and Breen, Watson, et al., 2010. See Wagner and Watson (2010) for further discussion. For some

examples of work on prosodic marking of syntactic boundaries in other languages, see Nespor and Vogel (1986/2007), Jun (2005b).

4. These examples are fluently produced utterances taken from data-sets the author has used for investigating prosody production and perception in conversational speech. (1a) is from the Buckeye corpus (Pitt et al., 2007), and (1b) is from the American MapTask Corpus, collected by Stefanie Shattuck-Hufnagel and shared with the author. The boundary transcriptions represent the consensus labelling of two or three independent, trained transcribers using the ToBI annotation system.
5. The studies employed different criteria for syntactic labelling, but both adopted an essentially clause-level analysis that identifies main clauses, relative clauses, parentheticals and clausal complements, and their internal structures, but which does not label multi-clause constituents as such.
6. Identifying pauses in speech is not entirely straightforward, since listeners' perceptions of pause do not necessarily coincide with silent intervals, but are also influenced by final lengthening and pitch (Nooteboom & Eefting, 1994). Moreover, the duration threshold for perceiving a silent interval as pause may depend on speech rate and other factors. Unfortunately, these concerns are not often addressed in works that report pause as an acoustic correlate of prosodic boundaries.
7. The terms 'topic unit' and 'topic structure' are used by authors who draw on the terminology of theories of discourse structure. For instance, Geluykens and Swerts (1994) use the term 'topic unit' as an informational segment, which in their speech materials is operationally defined in terms of the components of the linguistic task performed by their speakers. These authors note that topic units may overlap with talker turns: A single topic may continue across a change of turn, and on the other hand, a single turn may comprise more than one topic unit.
8. Ladd (2008, pp. 277–278) attributes the pattern of nuclear prominence on a sentence-final intransitive verb to prosodic phrasing: nuclear prominence on the predicate occurs only when the argument and predicate are in separate prosodic phrases. Ladd argues that in such structures the predicate may be perceived as having stronger prominence than the argument, which he takes as evidence for the recursive layering of prosodic phrase structure, with greater prominence assigned to the predicate's phrase, e.g.,[[DOGs]$_W$ [must be CARRIED]$_S$].
9. The status of F0 in the prosodic encoding of focus or information status is called to question by Kochanski, Grabe, Coleman, and Rosner (2005). In their corpus study of seven varieties of British English, they looked for acoustic correlates of prominence as marked by expert prosody transcribers. Kochanksi and colleagues find that among the many acoustic measures they tested, intensity and duration are correlated with prominence, but not measures of F0. Bearing in mind that prominent words do not necessarily express focus or new information (Calhoun, 2010a, 2010b), this finding leaves open the possibility that F0 does in fact mark focus or information status in their materials, but F0 may not be a reliable correlate of prominence construed more broadly to include prominence due to rhythm, accessibility or possibly other factors.
10. Another explanation for the perceived impoliteness of (9) might be that the focal prominence on *anything* evokes a set of alternatives ('you know [X]') which may be construed as impolite even without the rising intonation that Culpeper describes for this utterance. This analysis also serves to illustrate the point that prosody conveys a non-truth-conditional aspect of sentence meaning.
11. The authors do not identify the variety of English spoken by participants in their study, but they give institutional affiliations in South Africa, suggesting that the study reports on South African English.
12. Downstep patterns also arise in lexical tone systems, as widely reported for African and Chinese languages, where the second High tone in a sequence of High tones is realised with a step-wise lowered pitch. Downstep in lexical tone systems can be restricted to contexts where a Low tone intervenes between the successive High tones: H – L – H. For an overview of downstep in lexical tone systems and acoustic evidence for its interaction with phrase- and discourse-level prosody in Mandarin Chinese, see Wang and Xu (2011). Laniran and Clements (2003) present an acoustic study of downstep in the African language Yoruba. For an overview of downstep phenomena in a variety of African languages, see Hyman (2011) and references cited there.
13. Swerts and colleagues find that other properties related to pitch, such as range, register and contour shape and slope, also contribute to the perception of finality in utterances judged as appropriate at the end of a discourse unit, as noted in Section 3.3.
14. A central argument for a phonological representation of prosodic phrase structure lies in the frequent mismatch between prosodic and syntactic structure. Similarly, a phonological representation for prominence is motivated by the fact that prominence may correspond to semantic focus, information status (discourse-new or -given) or may be motivated on purely phonological grounds, to mark the beginning of a phrase or to promote rhythmic alternation across syllables in a phrase. See Ladd (2008, pp. 18–33) for further arguments supporting a phonological representation of prosody.

References

Adami, A. G., Mihaescu, R., Reynolds, D. A., & Godfrey, J. J. (2003). Modeling prosodic dynamics for speaker recognition. *Proceedings of Acoustics, Speech, and Signal Processing (ICASSP'03)*, *4*, 788–791.

Arnold, J. E. (2008). Reference production: Production-internal and addressee-oriented processes. *Language and Cognitive Processes*, *23*, 495–527. doi:10.1080/01690960801920099

Arvaniti, A., & Adamou, E. (2011). Focus expression in Romani. In M. B. Washburn, K. McKinney-Bock, E. Varis, A. Sawyer, & B. Tomaszewicz (Eds.), *Proceedings of the 28th West Coast Conference on Formal Linguistics* (pp. 240–248). Somerville, MA: Cascadilla Proceedings Project.

Arvaniti, A., Ladd, D. R., & Mennen, I. (2006). Phonetic effects of focus and 'tonal crowding' in intonation: Evidence from Greek polar questions. *Speech Communication*, *48*, 667–696. doi:10.1016/j.specom.2005.09.012

Avesani, C., Vayra, M., & Zmarich, C. (2007). On the articulatory bases of prominence in Italian. *Proceedings of the 16th International Congress on Phonetic Sciences*, 981–984.

Barth-Weingarten, D., Dehé, N., & Wichmann, A. (Eds.). (2009). *Where prosody meets pragmatics* (Studies in Pragmatics 8). Bingley: Emerald Group.

Baumann, S., & Grice, M. (2006). The intonation of accessibility. *Journal of Pragmatics*, *38*, 1636–1657. doi:10.1016/j.pragma.2005.03.017

Baumann, S., & Hadelich, K. (2003). Accent type and givenness: An experiment with auditory and visual priming. *Proceedings of the 15th International Congress of Phonetic Sciences*, 1811–1814.

Beckman, M. E. (1996). The parsing of prosody. *Language and Cognitive Processes*, *11*, 17–67. doi:10.1080/016909696387213

Beckman, M. E., & Edwards, J. (1990). Lengthenings and shortenings and the nature of prosodic constituency. In J. Kingston & M. E. Beckman (Eds.), *Papers in laboratory phonology I: Between the grammar and physics of speech* (pp. 152–178). Cambridge: Cambridge University Press.

Beckman, M. E., & Edwards, J. (1994). Articulatory evidence for differentiating stress categories. In P. A. Keating (Ed.), *Phonological structure and phonetic form: Papers in laboratory phonology III* (pp. 7–33). Cambridge: Cambridge University Press.

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

Beckman, M., & Pierrehumbert, J. (1986). Intonational structure in Japanese and English. *Phonology Yearbook*, *3*, 255–309. doi:10.1017/S095267570000066X

Birch, S., & Clifton, C., Jr., (1995). Focus, accent, and argument structure: Effects on language comprehension. *Language and Speech*, *38*, 365–392. doi:10.1177/002383099503800403

Blaauw, E. (1994). The contribution of prosodic boundary markers to the perceptual difference between read and spontaneous speech. *Speech Communication*, *14*, 359–375. doi:10.1016/0167-6393(94)90028-0

Bolinger, D. (1958). A theory of pitch accent in English. *Word*, *14*, 109–149.

Bolinger, D. (1982). Intonation and its parts. *Language*, *58*, 505–533. doi:10.2307/413847

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

Brazil, D. (1985). Phonology: Intonation in discourse. In D. van Teun (Ed.), *Handbook of discourse analysis, vol. 2. Dimensions of discourse* (pp. 57–75). London: Academic Press.

Breen, M., Fedorenko, E., Wagner, M., & Gibson, E. (2010). Acoustic correlates of information structure. *Language and Cognitive Processes*, *25*, 1044–1098. doi:10.1080/01690965.2010.504378

Breen, M., Watson, D. G., & Gibson, E. (2010). Intonational phrasing is constrained by meaning, not balance. *Language and Cognitive Processes*, *25*, 904–945. doi:10.1080/01690965.2010.508878

Brown, G. (1983). Prosodic structure and the given/new distinction. In A. Cutler & R. Ladd (Eds.), *Prosody: Models and measurements* (pp. 67–77). New York, NY: Springer.

Bryant, G., & Fox Tree, J. E. (2005). Is there an ironic tone of voice? *Language and Speech*, *48*, 257–277. doi:10.1177/00238309050480030101

Büring, D. (2006). Focus projection and default prominence. In V. Molnár & S. Winkler (Eds.), *The architecture of focus* (pp. 321–346). Berlin: Mouton De Gruyter.

Byrd, D., Kaun, A., Narayanan, S., & Saltzman, E. (2000). Phrasal signatures in articulation. In M. B. Broe & J. B. Pierrehumbert (Eds.), *Papers in laboratory phonology 5. Acquisition and the lexicon* (pp. 70–87). Cambridge: Cambridge University Press.

Byrd, D., Krivokapić, J., & Lee, S. (2006). How far, how long: On the temporal scope of prosodic boundary effects. *Journal of the Acoustical Society of America*, *120*, 1589–1599. doi:10.1121/1.2217135

Byrd, D., & Saltzman, E. (1998). Intragestural dynamics of multiple phrasal boundaries. *Journal of Phonetics*, *26*, 173–199. doi:10.1006/jpho.1998.0071

Calhoun, S. (2006). *Information structure and the prosodic structure of English: A probabilistic relationship* (Unpublished doctoral dissertation). University of Edinburgh, Edinburgh, United Kingdom.

Calhoun, S. (2010a). The centrality of metrical structure in signaling information structure: A probabilistic perspective. *Language*, *86*, 1–42. doi:10.1353/lan.0.0197

Calhoun, S. (2010b). How does informativeness affect prosodic prominence? *Language and Cognitive Processes*, *25*, 1099–1140. doi:10.1080/01690965.2010.491682

Cambier-Langeveld, T. (1997). The domain of final lengthening in the production of Dutch. In H. de Hoop & J. Coerts (Eds.), *Linguistics in the Netherlands* (pp. 13–24). Amsterdam: John Benjamins.

Cambier-Langeveld, T. (1999). The interaction between final lengthening and accentual lengthening: Dutch versus English. In J. J. Ohala (Ed.), *Proceedings of the 14th International Congress of Phonetic Sciences, vol. 1* (pp. 467–470). San Francisco, CA: University of California, Berkeley.

Cambier-Langeveld, T., & Turk, A. (1999). A cross-linguistic study of accentual lengthening: Dutch vs. English. *Journal of Phonetics*, *27*, 255–280. doi:10.1006/jpho.1999.0096

Campbell, N., & Beckman, M. (1997). Stress, prominence, and spectral tilt. In *Proceedings of INT-1997*, 67–70. Retrieved September 9, 2012, from http://www.isca-speech.org/archive_open/int_97/inta_067.html

Cao, J. (2004). Restudy of segmental lengthening in Mandarin Chinese. In B. Bel & I. Marlien (Eds.), *Proceedings of Speech Prosody* (pp. 231–234). Nara, Japan.

Carlson, K., Clifton, C., Jr., & Frazier, L. (2001). Prosodic boundaries in adjunct attachment. *Journal of Memory and Language*, *45*, 58–81. doi:10.1006/jmla.2000.2762

Caspers, J. (2003). Local speech melody as a limiting factor in the turn-taking system in Dutch. *Journal of Phonetics*, *31*, 251–276. doi:10.1016/S0095-4470(03)00007-X

Chafe, W. (1987). Cognitive constraints on information flow. In R. Tomlin (Ed.), *Coherence and grounding in discourse* (pp. 20–51). Amsterdam: John Benjamins.

Cheang, H., & Pell, M. (2008). The sound of sarcasm. *Speech Communication*, *50*, 366–381. doi:10.1016/j.specom.2007.11.003

Chen, A., den Os, E., & de Ruiter, J. P. (2007). Pitch accent type matters for online processing of information status: Evidence from natural and synthetic speech. *The Linguistic Review*, *24*, 317–344. doi:10.1515/TLR.2007.012

Chen, A., Gussenhoven, C., & Rietveld, T. (2004). Language-specificity in the perception of paralinguistic intonational Meaning. *Language and Speech*, *47*, 311–349. doi:10.1177/00238309040470040101

Cho, T. (2005). Prosodic strengthening and featural enhancement: Evidence from acoustic and articulatory realizations of /ɑ,i/ in English. *Journal of the Acoustical Society of America*, *117*, 3867–3878. doi:10.1121/1.1861893

Cho, T., & Keating, P. (2001). Articulatory and acoustic studies of domain-initial strengthening in Korean. *Journal of Phonetics*, *29*, 155–190. doi:10.006/jpho.2001.0131

Cho, T., & Keating, P. (2009). Effects of initial position versus prominence in English. *Journal of Phonetics*, *37*, 466–485. doi:10.1016/j.wocn.2009.08.001

Cho, T., & McQueen, J. M. (2005). Prosodic influences on consonant production in Dutch: Effects of prosodic boundaries, phrasal accent and lexical stress. *Journal of Phonetics*, *33*, 121–157. doi:10.1016/j.wocn.2005.01.001

Clifton, C., Carlson, K., & Frazier, L. (2002). Informative prosodic boundaries. *Language and Speech*, *45*, 87–114. doi:10.1177/00238309020450020101

Cole, J., & Hasegawa-Johnson, M. (2012). Corpus phonology with speech resources. In A. Cohn, C. Fougeron, & M. Huffman (Eds.), *Handbook of laboratory phonology* (pp. 431–440). Oxford: Oxford University Press.

Cole, J., Hasegawa-Johnson, M., Shih, C., Kim, H., Lee, E., Lu, H., … Yoon, T. (2005). Prosodic parallelism as a cue to repetition disfluency. In J. Véronis & E. Campione (Eds.), *Proceedings of DiSS '05: Disfluency in Spontaneous Speech Workshop* (pp. 53–58). Aix-en-Provence, France.

Cole, J., Kim, H., Choi, H., & Hasegawa-Johnson, M. (2007). Prosodic effects on acoustic cues to stop voicing and place of articulation: Evidence from radio news speech. *Journal of Phonetics*, *35*, 180–209. doi:10.1016/j.wocn.2006.03.004

Cole, J., McMurray, B., Linebaugh, G., & Munson, C. (2010). Unmasking the acoustic effects of vowel-to-vowel coarticulation: A statistical modeling approach. *Journal of Phonetics*, *38*, 167–184. doi:10.1016/j.wocn.2009.08.004

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*, 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*, 425–452. doi:10.1515/labphon.2010.022

Cole, J., & Shattuck-Hufnagel, S. (2011). The phonology and phonetics of perceived prosody: What do listeners imitate? In P. Cosi, R. De Mori, G. Di Fabbrizio, & R. Pieraccini (Eds.), *Proceedings of Interspeech* (pp. 969–972). Florence, Italy.

Cooper, W. E., Eady, S. J., & Mueller, P. R. (1985). Acoustical aspects of contrastive stress in question-answer contexts. *Journal of the Acoustical Society of America*, *77*, 2142–2156. doi:10.1121/1.392372

Couper-Kuhlen, E. (1996). The prosody of repetition: On quoting and mimicry. In E. Couper-Kuhlen & M. Selting (Eds.), *Prosody in conversation: Interactional studies* (pp. 366–405). Cambridge: Cambridge University Press.

Cruttenden, A. (1986/1997). *Intonation*. Cambridge: Cambridge University Press.

Culpeper, J. (2011). "It's not what you said, it's how you said it!": Prosody and impoliteness. In Linguistic Politeness Research Group (Eds.), *Discursive approaches to politeness* (pp. 57–83). Berlin: De Gruyter Mouton.

Culpeper, J., Bousfield, D., & Wichmann, A. (2003). Impoliteness revisited: With special reference to dynamic and prosodic aspects. *Journal of Pragmatics*, *35*, 1545–1579. doi:10.1016/S0378-2166(02)00118-2

Dahan, D., Tanenhaus, M. K., & Chambers, C. G. (2002). Accent and reference resolution in spoken-language comprehension. *Journal of Memory and Language*, *47*, 292–314. doi:10.1016/S0749-596X(02)00001-3

Daly, N., & Warren, P. (2001). Pitching it differently in New Zealand English: Speaker sex and intonation. *Journal of Sociolinguistics*, *5*, 85–96. doi:10.1111/1467-9481.00139

de Jong, K. (1995). The supraglottal articulation of prominence in English: Linguistic stress as localized hyperarticulation. *Journal of the Acoustical Society of America*, *97*, 491–504. doi:10.1121/1.412275

de Jong, K., Beckman, M. E., & Edwards, J. (1993). The interplay between prosodic structure and coarticulation. *Language and Speech 36*, 197–212. doi:10.1177/002383099303600305

de Jong, K., & Zawaydeh, B. (1999). Stress, duration, and intonation in Arabic word-level prosody. *Journal of Phonetics*, *27*, 3–22. doi:10.006/jpho.2001.0151

de Jong, K., & Zawaydeh, B. (2002). Comparing stress, lexical focus, and segmental focus: Patterns of variation in Arabic vowel duration. *Journal of Phonetics*, *30*, 53–75. doi:10.1006/jpho.2001.0151

Delais-Roussarie, E., & Rialland, A. (2007). Metrical structure, tonal association and focus in French. In S. Baauw (Ed.), *Romance languages and linguistic theory 2005: Selected papers from going romance Utrecht* (pp. 73–98). Amsterdam: John Benjamins.

den Ouden, H., Noordman, L., & Terken, J. (2009). Prosodic realizations of global and local structure and rhetorical relations in read aloud news reports. *Speech Communication*, *51*, 116–129. doi:10.1016/j.specom.2008.06.003

de Pijper, J. R., & Sanderman, A. A. (1994). On the perceptual strength of prosodic boundaries and its relation to suprasegmental cues. *Journal of the Acoustical Society of America*, *96*, 2037–2047. doi:10.1121/1.410145

Dilley, L., Shattuck-Hufnagel, S., & Ostendorf, M. (1996). Glottalization of word-initial vowels as a function of prosodic structure. *Journal of Phonetics*, *24*, 423–444. doi:10.1006/jpho.1996.0023

D'Imperio, M., & House, D. (1997). Perception of questions and statements in Neapolitan Italian. In G. Kokkinakis, N. Fakotakis, & E. Dermatas (Eds.), *Proceedings of Eurospeech*, 251–254.

Dogil, G., & Williams, B. (1999). The phonetic manifestation of word stress. In H. van der Hulst (Ed.), *Word prosodic systems in the languages of Europe* (pp. 273–311). Berlin: Mouton de Gruyter.

Donati, C., & Nespor, N. (2003). From focus to syntax. *Lingua*, *113*, 1119–1142. doi:10.1016/S0024-3841(03)00015-9

Downing, L. (2011). Bantu tone. In M. van Oostendorp, C. J. Ewen, E. Hume, & K. Rice (Eds.), *The Blackwell companion to phonology*. Blackwell. Retrieved from http://www.companiontophonology.com

Eady, S., & Cooper, W. (1986). Speech intonation and focus location in matched statements and questions. *Journal of the Acoustical Society of America*, *80*, 402–415. doi:10.1121/1.394091

Eady, S., Cooper, W., Klouda, G., Mueller, P., & Lotts, D. (1986). Acoustical characterization of sentential focus: Narrow vs. broad and single vs. dual focus environments. *Language and Speech*, *29*, 233–250. doi:10.1177/002383098602900304

Edlund, J., & Heldner, M. (2005). Exploring prosody in interaction control. *Phonetica*, *62*, 215–226. doi:10.1159/000090099

Edlund, J., House, D., & Strömbergsson, S. (2012). Question types and some prosodic correlates in 600 questions in the Spontal database of Swedish dialogues. In Q. Ma, H. Ding, & D. Hirst (Eds.), *Proceedings of Speech Prosody* (pp. 737–740). Shanghai.

Edwards, J., Beckman, M. E., & Fletcher, J. (1991). The articulatory kinematics of final lengthening. *Journal of the Acoustical Society of America*, *89*, 369–382. doi:10.1121/1.400674

Face, T. L. (2005). F0 peak height and the perception of sentence type in Castilian Spanish. *Revista internacional de lingüística iberoamericana*, *3*(2), 49–65.

Farahani, F., Georgiou, P. G., & Narayanan, S. S. (2004). Speaker identification using supra-segmental pitch pattern dynamics. *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing-ICASSP*, 89–92. doi:10.1109/ICASSP.2004.1325929

Ferý, C. (1993). *German intonational patterns*. Tübingen: Niemeyer.

Ferý, C., & Kügler, F. (2008). Pitch accent scaling on given, new and focused constituents in German. *Journal of Phonetics*, *36*, 680–703. doi:10.1016/j.wocn.2008.05.001

Fougeron, C. (2001). Articulatory properties of initial segments in several prosodic constituents in French. *Journal of Phonetics*, *29*, 109–135. doi:10.1006/jpho.2000.0114

Fougeron, C., & Keating, P. A. (1997). Articulatory strengthening at edges of prosodic domains. *Journal of the Acoustical Society of America*, *106*, 3728–3740. doi:10.1121/1.418332

Fowler, C. A., & Brown, J. M. (2000). Perceptual parsing of acoustic consequences of velum lowering from information for vowels. *Perception & Psychophysics*, *62*, 21–32. doi:10.3758/BF03212058

Fowler, C. A., & Smith, M. R. (1986). Speech perception as "vector analysis": An approach to the problems of segmentation and invariance. In J. Perkell & D. Klatt (Eds.), *Invariance and variability in speech processes* (pp. 123–136). Hillsdale, NJ: Erlbaum.

Frazier, L., Clifton, C., Jr., & Carlson, K. (2004). Don't break or do: Prosodic boundary preferences. *Lingua*, *114*, 3–27. doi:10.1016/S0024-3841(03)00044-5

Frota, S. (2000). *Prosody and focus in European Portuguese: Phonological phrasing and intonation* (Outstanding dissertations in linguistics). Garland Press, New York.

Fry, D. B. (1955). Duration and intensity as physical correlates of linguistic stress. *Journal of the Acoustical Society of America*, *27*, 765–768. doi:10.1121/1.1908022

Geluykens, R., & Swerts, M. (1994). Prosodic cues to discourse boundaries in experimental dialogues. *Speech Communication*, *15*, 69–77. doi:10.1016/0167-6393(94)90042-6

Godfrey, J., Holliman, E., & McDaniel, J. (1992). SWITCHBOARD: Telephone speech corpus for research and development. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP-92* (pp. 517–520). San Francisco, CA.

Gordon, M. (2011). Stress: Phonotactic and phonetic evidence. In M. van Oostendorp, C. J. Ewen, E. Hume, & K. Rice (Eds.), *The Blackwell companion to phonology*. Blackwell. Retrieved from http://www.companiontophonology.com

Grabe, E. (2004). Intonational variation in urban dialects of English spoken in the British Isles. In P. Gilles & J. Peters (Eds.), *Regional variation in intonation* (pp. 9–31). Tuebingen: Niemeyer.

Grabe, E., Gussenhoven, C., Haan, J., Marsi, E., & Post, B. (1997). The meaning of intonation phrase onsets in Dutch. In A. Botinis, G. Kouroupetroglou, & G. Carayiannis (Eds.), *Intonation: Theory, models, and applications* (pp. 161–164). Athens: ESCA. Graff, Delia.

Grabe, E., Post, B., Nolan, F., & Farrar, K. (2000). Pitch accent realization in four varieties of British English. *Journal of Phonetics*, *28*, 161–185. doi:10.006/jpho.2000.0111

Gravano, A., Hirschberg, J., & Beňuš,Š. (2012). Affirmative cue words in task-oriented dialogue. *Computational Linguistics*, *38*(1), 1–39. doi:10.1162/COLI_a_00083

Grice, M. (1995). *The intonation of interrogation in Palermo Italian: Implications for intonation theory.* Tübingen: Niemeyer.

Grice, M., Reyelt, M., Benzmuller, R., Mayer, J., & Batliner, A. (1996). Consistency in transcription and labelling of German intonation with GToBI. *Proceedings of the International Conference on Spoken Language Processing*, *3*, 1716–1719.

Grosz, B., & Hirschberg, J. (1992). Some intonational characteristics of discourse structure. In *International Conference on Spoken Language Processing- ICSLP 92* (pp. 429–432). Banff, Alberta.

Gussenhoven, C. (1984). *On the grammar and semantics of sentence accents*. Dordrecht: Foris.

Gussenhoven, C. (2002). Intonation and interpretation: Phonetics and phonology. In *Speech Prosody* (pp. 47–57). Aix-en-Provence.

Gussenhoven, C. (2004). *The phonology of tone and intonation*. Cambridge: Cambridge University Press.

Gussenhoven, C., Repp, B. H., Rietveld, A., Rump, H. H., & Terken, J. (1997). The perceptual prominence of fundamental frequency peaks. *Journal of the Acoustical Society of America*, *102*, 3009–3022. doi:10.1121/1.420355

Gussenhoven, C., & Rietveld, A. C. (1988). Fundamental frequency declination in Dutch: Testing three hypotheses. *Journal of Phonetics*, *16*, 355–369.

Haan, J., & van Heuven, V. J. (1999). Male vs. female pitch range in Dutch questions. In *Proceedings of the Thirteenth International Congress of Phonetic Sciences* (pp. 1581–1584). San Francisco, CA.

Halliday, M. A. K. (1967). *Intonation and grammar in British English*. The Hague: Mouton.

Hammond, M. (2011). The foot. In M. van Oostendorp, C. J. Ewen, E. Hume, & K. Rice (Eds.), *The Blackwell companion to phonology*. Blackwell. Retrieved from http://www.companiontophonology.com

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

Hazan, V., & Baker, R. (2010). Does reading clearly produce the same acoustic-phonetic modifications as spontaneous speech in a clear speaking style? In *Proceedings of DiSS-LPSS Joint Workshop* (pp. 7–10). Tokyo.

Heldner, M., & Edlund, J. (2010). Pauses, gaps and overlaps in conversations. *Journal of Phonetics*, *38*, 555–568. doi:10.1016/j.wocn.2010.08.002

Heldner, M., Edlund, J., & Hirschberg, J. (2010). Pitch similarity in the vicinity of backchannels. In T. Kobayashi, K. Hirose, & S. Nakamura (Eds.), *Proceedings of Interspeech* (pp. 3054–3057). Makuhari.

Herman, R. 2000. Phonetic markers of global discourse structures in English. *Journal of Phonetics*, *28*, 466–493. doi:10.1006/jpho.2000.0127

Hirschberg, J. (2000). A corpus-based approach to the study of speaking style. In M. Horne (Ed.), *Prosody: Theory and experiment. Studies presented to Gösta Bruce* (pp. 335–350). Dordrecht: Kluwer Academic.

Hirschberg, J. (2002). Communication and prosody: Functional aspects of prosody. *Speech Communication*, *36*, 31–43. doi:10.1016/S0167-6393(01)00024-3

Hirst, D., & Bouzon, C. (2005). The effect of stress and boundaries on segmental duration in a corpus of authentic speech (British English). In *Proceedings of Interspeech* (pp. 29–32). Lisbon.

Hirst, D., di Cristo, A., & Espesser, R. (2000). Levels of representation and levels of analysis for the description of intonation systems. In M. Horne (Ed.), *Prosody: Theory and experiment. Studies presented to Gösta Bruce* (pp. 51–88). Dordrecht: Kluwer.

Hockey, B. A., & Fagyal, Z. (1999). Phonemic length and preboundary lengthening: An experimental investigation on the use of durational cues in Hungarian. In *Proceedings of the XIVth International Congress of Phonetic Sciences* (pp. 313–316). San Francisco, CA.

Horne, M., Strangert, E., & Heldner, M. (1995). Prosodic boundary strength in Swedish: Final lengthening and silent interval duration. In T. Kobayashi, K. Hirose, & S. Nakamura (Eds.), *Proceedings of the 13th International Congress of Phonetic Sciences* (pp. 170–173). Stockholm.

House, D. (2003). Perceiving question intonation: The role of pre-focal pause and delayed focal peak. In *Proceedings of the 15th International Congress of Phonetic Sciences* (pp. 755–758). Barcelona.

Hualde, J. I. (2000). Intonation in Spanish and the other Ibero-Romance languages: Over-view and status quaestionis. In C. Wiltshire & J. Camps (Eds.), *Romance phonology and variation: Selected papers from LSRL 30* (pp. 101–115). Amsterdam: John Benjamins.

Hyman, L. M. (2011). The representation of tone. In M. van Oostendorp, C. J. Ewen, E. Hume, & K. Rice (Eds.), *The Blackwell companion to phonology.* Blackwell. Retrieved from http://www.companiontophonology.com

Ito, K., & Speer, S. R. (2008). Anticipatory effects of intonation: Eye movements during instructed visual search. *Journal of Memory and Language*, *85*, 541–573. doi:10.1016/j.jml.2007.06.013

Jakobson, R., Fant, G., & Halle, M. (1951). *Preliminaries to speech analysis: The distinctive features and their correlates*. Cambridge, MA: MIT Press.

Jun, S. A. (2005a). Korean intonational phonology and prosodic transcription. In S. Jun (Ed.), *Prosodic typology: The phonology of intonation and phrasing* (pp. 201–229). Oxford: Oxford University Press.

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

Kager, R. (1995). The metrical theory of word stress. In J. Goldsmith (Ed.), *The handbook of phonological theory* (pp. 367–402). Oxford: Blackwell.

Katz, J., & Selkirk, E. (2011). Contrastive focus vs. discourse-new: Evidence from phonetic prominence in English. *Language*, *87*, 771–816. doi:10.1353/lan.2011.0076

Keating, P., Cho, T., Fougeron, C., & Hsu, C. (2003). Domain-initial strengthening in four languages. In J. Local, R. Ogden, & R. Temple (Eds.), *Papers in laboratory phonology VI* (pp. 143–161). Cambridge: Cambridge University Press.

Klatt, D. H. (1976). Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *Journal of the Acoustical Society of America*, *59*, 1208–1221. doi:10.1121/1.380986

Klewitz, G., & Couper-Kuhlen, E. (1999). Quote-unquote: The role of prosody in the contextualization of reported speech sequences. *Pragmatics: Quarterly Publication of the International Pragmatics Association*, *9*, 459–485.

Kochanski, G., Grabe, E., Coleman, J., & Rosner, B. (2005). Loudness predicts prominence: Fundamental frequency lends little. *Journal of the Acoustical Society of America*, *118*, 1038–1054. doi:10.1121/1.1923349

Krahmer, E., & Swerts, M. (2001). On the alleged existence of contrastive accents. *Speech Communication*, *34*, 391–405. doi:10.1016/S0167-6393(00)00058-3

Kraljic, T., & Brennan, S. E. (2005). Prosodic disambiguation of syntactic structure: For the speaker or for the addressee? *Cognitive Psychology*, *50*, 194–231.doi:10.1016/j.cogpsych.2004.08.002

Krivokapić, J., & Byrd, D. (2012). Prosodic boundary strength: An articulatory and perceptual study. *Journal of Phonetics*, *40*, 430–442. doi:10.1016/j.wocn.2012.02.011

Ladd, D. R. (1988). Declination “reset” and the hierarchical organization of utterances. *The Journal of the Acoustical Society of America*, *84*, 530–544. doi:10.1121/1.396830

Ladd, D. R. (2008). *Intonational phonology* (2nd ed.). Cambridge and New York, NY: Cambridge University Press.

Ladd, D. R., & Morton, R. (1997). The perception of intonational emphasis: Continuous or categorical? *Journal of Phonetics*, *25*, 313–342. doi:10.1006/jpho.1997.0046

Ladd, D. R., Silverman, K. E., Tolkmitt, F., Bergmann, G., & Scherer, K. R. (1985). Evidence for the independent function of intonation contour type, voice quality, and F0 range in signaling speaker affect. *Journal of the Acoustical Society of America*, *78*, 435–444. doi:10.1121/1.392466

Laniran, Y. O., & Clements, G. N. (2003). Downstep and high raising: Interacting factors in Yoruba tone production. *Journal of Phonetics*, *31*, 203–250. doi:10.1016/S0095-4470(02)00098-0

Laskowski, K., Heldner, M., & Edlung, J. (2009). Exploring the prosody of floor mechanisms in English using the fundamental frequency variation spectrum. In *Proceedings of the 17th European Signal Processing Conference (EUSIPCO 2009)* (pp. 2539–2543). Glasgow.

Lehiste, I. (1970). *Suprasegmentals*. Cambridge, MA: MIT Press.

Levitan, R., Gravano, A., & Hirschberg, J. (2011). Entrainment in speech preceding backchannels. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics (short papers)* (pp. 113–117).

Levitan, R., Gravano, A., Willson, L., Benuš, Š., Hirschberg, J., & Nenkova, A. (2012). Acoustic-prosodic entrainment and social behavior. In *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human language technologies* (pp. 11–19).

Levitan, R., & Hirschberg, J. (2011). Measuring acoustic-prosodic entrainment with respect to multiple levels and dimensions. In P. Cosi, R. De Mori, G. Di Fabbrizio, & R. Pieraccini (Eds.), *Proceedings of Interspeech* (pp. 3081–3084). Florence, Italy.

Liberman, M. (1975). *The intonational system of English* (Unpublished doctoral dissertation). Cambridge, MA: Massachusetts Institute of Technology. Distributed by Indiana University Linguistics Club (1978).

Liberman, M., & Prince, A. (1977). On stress and linguistic rhythm. *Linguistic Inquiry*, *8*, 249–336. Retrieved from http://www.jstor.org/stable/4177987

Lieberman, P. (1960). Some acoustic correlates of word stress in American English. *Journal of the Acoustical Society of America*, *32*, 451–454. doi:10.1121/1.1908095

Lindblom, B. E. F. (1990). Explaining phonetic variation: A sketch of the H & H theory. In H. J. Hardcastle & A. Marchal (Eds.), *Speech production and speech modeling* (pp. 403–439). *NATO ASI Seriess D: Behaviour and Social Sciences*, *55*. Dordrecht: Kluwer A.P.

Local, J., Kelly, J., & Wells, W. (1986). Towards a phonology of conversation: Turn taking in Tyneside English. *Journal of Linguistics*, *22*, 411–437. doi:10.1017/S0022226700010859

Low, E. L., Grabe, E., & Nolan, F. (2001). Quantitative characterisations of speech rhythm: Syllable-timing in Singapore English. *Language and Speech*, *43*, 377–401. doi:10.1177/00238309000430040301

Luchkina, T., & Cole, J. (2013). Routes to prominence in free word order language discourse. In P. Mertens & A. C. Simon (Eds.), *Proceedings of the Prosody-Discourse Interface Conference 2013 (IDP-2013)*. Leuven. http://www.arts.kuleuven.be/ling/franitalco/idp2013/proceedings.html.

Luchkina, T., & Cole, J. (2014). Structural and prosodic correlates of prominence in free word order language discourse. In N. Campbell, D. Gibbon, & D. Hirst (Eds.), *Proceedings of Speech Prosody 7*. Dublin.

Magne, C., Astésano, C., Lacheret-Dujour, A., Morel, M., Alter, K., & Besson, M. (2005). On-line processing of “pop-out” words in spoken French dialogues. *Journal of Cognitive Neuroscience*, *17*, 740–756. doi:10.1162/0898929053747667

Mattys, S., White, L., & Melhorn, J. (2005.) Integration of multiple speech segmentation cues: A hierarchical framework. *Journal of Experimental Psychology*, *134*, 477–500. doi:10.1037/0096-3445.134.4.477

McMurray, B., Cole, J., & Munson, C. (2011). Features as an emergent product of perceptual parsing: Evidence from vowel-to-vowel coarticulation. In C. N. Clements & R. Ridouane (Eds.), *Where do phonological features come from? Cognitive, physical and developmental bases of distinctive speech categories* (pp. 197–236). Amsterdam: John Benjamins.

McMurray, B., & Jongman, A. (2011). What information is necessary for speech categorization? Harnessing variability in the speech signal by integrating cues computed relative to expectations. *Psychological Review*, *118*, 219–246. doi:10.1037/a0022325

Menn, L., & Boyce, S. (1982). Fundamental frequency and discourse structure. *Language and Speech*, *25*, 341–383. doi:10.1177/002383098202500403

Mo, Y. (2011). *Prosody production and perception with conversational speech* (Doctoral dissertation). Retrieved from http://www.ideals.illinois.edu/handle/2142/18560

Mooshammer, C., Bombien, L., & Krivokapic, J. (2013). Prosodic effects on speech gestures: A shape analysis based on functional data analysis. *Proceedings of Meetings on Acoustics*, *19*, 060182. Acoustical Society of America. doi:10.1121/1.4800316

Mooshammer, C., & Fuchs, S. (2002). Stress distinction in German: Simulating kinematic parameters of tongue-tip gestures. *Journal of Phonetics*, *30*, 337–355. doi:10.1006/jpho.2001.0159

Nakajima, S., & Allen, J. F. (1993). A study on prosody and discourse structure in cooperative dialogues. *Phonetica*, *50*, 197–210. doi:10.1159/000261940

Nakatani, C. H., & Hirschberg, J. (1994). A corpus-based study of repair cues in spontaneous speech. *Journal of the Acoustical Society of America*, *95*, 1603–1616. doi:10.1121/1.408547

Nespor, M., & Vogel, I. (1986/2007). *Prosodic phonology*. Dordrecht: Foris.Berlin: Mouton de Gruyter.

Nooteboom, S. G., & Eefting, W. (1994). Evidence for the adaptive nature of speech on the phrase level and below. *Phonetica*, *51*, 92–98. doi:10.1159/000261961

Ohala, J. J. (1983). Cross-language use of pitch: An ethological view. *Phonetica*, *40*, 1–18. doi:10.1159/000261678

Ohala, J. J. (1984). An ethological perspective on common cross-language utilization of F0 in voice. *Phonetica*, *41*, 1–16. doi:10.1159/000261706

Oliveira, M., & Freitas, T. (2008). Intonation as a cue to turn management in telephone and face-to-face interactions. In P. A. Barbosa, S. Madureira, & C. Reis (Eds.), *Proceedings of Speech Prosody* (pp. 485–488). Campiñas.

Ortega-Llebaria, M., & Prieto, P. (2011). Acoustic correlates of stress in central Catalan and Castilian Spanish. *Language and Speech*, *54*(1), 73–97. doi:10.1177/0023830910388014

Pell, M., Paulmann, S., Dara, C., Alasseri, A., & Kotz, S. (2009). Factors in the recognition of vocally expressed emotions: A comparison of four languages. *Journal of Phonetics*, *37*, 417–435. doi:10.1016/j.wocn.2009.07.005

Peppé, S., Wells, J., & Maxim, B. (2000). Prosodic variation in Southern British English. *Language and Speech*, *43*, 309–334. doi:10.1177/00238309000430030501

Pierrehumbert, J. (1979). The perception of fundamental frequency declination. *Journal of the Acoustical Society of America*, *66*, 363–369. doi:10.1121/1.383670

Pierrehumbert, J. (1980). *The phonology and phonetics of English intonation* (Doctoral dissertation). Cambridge, MA: Massachusetts Institute of Technology. Distributed by the Indiana University Linguistics Club (1988).

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

Pitt, M. A., Dilley, L., Johnson, K., Kiesling, S., Raymond, W., Hume, E., & Fosler-Lussier, E. (2007). *Buckeye corpus of conversational speech* (2nd release). Columbus, OH: Department of Psychology, Ohio State University. Retrieved March 15, 2006, from www.buckeyecorpus.osu.edu

Pitrelli, J. F., Beckman, M. E., & Hirschberg, J. (1994). Evaluation of prosodic transcription labeling reliability in the ToBI framework. In *Proceedings of the International Conference on Spoken Language Processing* (pp. 123–126). Yokohama.

Plag, I., Kunter, G., & Schramm, M. (2011.) Acoustic correlates of primary and secondary stress in North American English. *Journal of Phonetics*, *39*, 362–374. doi:10.1016/j.wocn.2011.03.004

Porges, S. W. 2011. *The polyvagal theory: Neurophysiological foundations of emotions, attachment, communication, and self-regulation*. New York, NY: WW Norton.

Porges, S. W., Macellaio, M., Stanfill, S. D., McCue, K., Lewis, G. F., Harden, E. R., … Heilman, K. J. (2013). Respiratory sinus arrhythmia and auditory processing in autism: Modifiable deficits of an integrated social engagement system? *International Journal of Psychophysiology*, *88*, 261–270. doi:10.1016/j.ijpsycho.2012.11.009

Prieto, P., van Santen, J., & Hirshberg, J. (1995). Tonal alignment patterns in Spanish. *Journal of Phonetics*, *23*, 429–451. doi:10.1006/jpho.1995.0032

Redi, L., & Shattuck-Hufnagel, S. (2001). Variation in the realization of glottalization in normal speakers. *Journal of Phonetics*, *29*, 407–429. doi:10.1006/jpho.2001.0145

Rooth, M. (1992). A theory of focus interpretation. *Natural Language Semantics*, *1*, 75–116. doi:10.1007/BF02342617

Rump, H. H., & Collier, R. (1996). Focus conditions and the prominence of pitch-accented syllables. *Language and Speech*, *39*(1), 1–17. doi:10.1177/002383099603900101

Sanderman, A. A., & Collier, R. (1997). Prosodic phrasing and comprehension. *Language and Speech*, *40*, 391–409. doi:10.1177/002383099704000405

Schafer, A., Speer, S., Warren, P., & White, S. D. (2000). Intonational disambiguation in sentence production and comprehension. *Journal of Psycholinguistic Research*, *29*, 169–182. doi:10.1023/A:1005192911512

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

Selkirk, E. (1986). On derived domains in sentence phonology. *Phonology Yearbook*, *3*, 371–375. doi:10.1017/S0952675700000695

Selkirk, E. (1995). Sentence prosody: Intonation, stress and phrasing. In J. A. Goldsmith (Ed.), *The handbook of phonological theory* (pp. 550–569). Cambridge, MA: Blackwell.

Selting, M. (2000). The construction of units in conversational talk. *Language in Society*, *29*, 477–517. doi:10.1017/s0047404500004012

Shattuck-Hufnagel, S., Ostendorf, M., & Ross, K. (1994). Stress shift and early pitch accent placement in lexical items in American English. *Journal of Phonetics*, *22*, 357–388.

Shenk, P. S. (2006). The interactional and syntactic importance of prosody in Spanish-English bilingual discourse. *International Journal of Bilingualism*, *10*, 179–205. doi:10.1177/13670069060100020401

Shriberg, E. 2001. To 'errrr' is human: Ecology and acoustics of speech disfluencies. *Journal of the International Phonetic Association*, *31*, 153–164. doi:10.1017/S0025100301001128

Silverman, K. E. A., & Pierrehumbert, J. B. (1990). The timing of prenuclear high accents in English. In J. Kingston & M. E. Beckman (Eds.), *Papers in laboratory phonology 1: Between the grammar and the physics of speech* (pp. 72–106). Cambridge: Cambridge University Press.

Skopeteas, S., & Fanselow, G. (2010). Focus in Georgian and the expression of contrast. *Lingua*, *120*, 1370–1391. doi:10.1016/j.lingua.2008.10.012

Slioussar, N. (2011). *Grammar and information structure: A novel view based on Russian data*. Retrieved from September 13, 2013, from http://www.slioussar.ru/talks-papers.html

Sluijter, A. M. C., & Terken, J. M. B. (1993). Beyond sentence prosody: Paragraph intonation in Dutch. *Phonetica*, *50*, 180–188. doi:10.1159/000261938

Sluijter, A. M. C., & van Heuven, V. (1996a). Spectral tilt as an acoustic correlate of linguistic stress. *Journal of the Acoustical Society of America*, *100*, 2471–2485. doi:10.1121/1.417955

Sluijter, A. M. C., & van Heuven, V. (1996b). Acoustic correlates of linguistic stress and accent in Dutch and American English. In *Proceedings of International Conference on Spoken Language Processing-ICSLP 96* (pp. 630–633). Philadelphia, PA: Applied Science and Engineering Laboratories, Alfred I. duPont Institute.

Smith, C. L. (2004).Topic transitions and durational prosody in reading aloud: Production and modeling. *Speech Communication*, *42*, 247–270. doi:10.1016/j.specom.2003.09.004

Snedeker, J., & Truesell, 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

Sridhar, V. K. R., Bangalore, S., & Narayanan, S. (2009). Combining lexical, syntactic and prosodic cues for improved online dialog act tagging. *Computer Speech and Language*, *23*, 407–422. doi:10.1016/j.csl.2008.12.001

Swerts, M. (1997). Prosodic features at discourse boundaries of different strength. *Journal of the Acoustical Society of America*, *101*, 514–521. doi:10.1121/1.418114

Swerts, M., Bouwhuis, D. G., & Collier, R. (1994). Melodic cues to the perceived "finality" of utterances. *Journal of the Acoustical Society of America*, *96*, 2064–2075. doi:10.1121/1.410148

Swerts, M., & Geluykens, R. (1993). The prosody of information units in spontaneous monologue. *Phonetica*, *51*, 189–196. doi:10.1159/000261939

Swerts, M., & Geluykens, R. (1994). Prosody as a marker of information flow in spoken discourse. *Language and Speech*, *37*, 21–43. doi:10.1177/002383099403700102

Swerts, M., & Hirschberg, J. (2008). Prosodic predictors of upcoming positive or negative content in spoken messages. *Journal of the Acoustical Society of America*, *128*, 1337–1344. doi:10.1121/1.3466875

Swerts, M., Krahmer, E., & Avesani, C. (2002). Prosodic marking of information status in Dutch and Italian: A comparative analysis. *Journal of Phonetics*, *30*, 629–654. doi:10.1006/jpho.2002.0178

Swerts, M., Strangert, E., & Heldner, M. (1996). F0 declination in read-aloud and spontaneous speech. *Proceedings of the International Conference on Spoken Language Processing (ICSLP)*, 1501–1504. doi:10.1109/ICSLP.1996.607901

Terken, J., & Nooteboom, S. G. (1987) Opposite effects of accentuation and deaccentuation on verification latencies for given and new information. *Language and Cognitive Processes*, *2*, 3–4, 145–163. doi:10.1080/01690968708406928

t'Hart, J., Collier, R., & Cohen, A. (1990). *A perceptual study of intonation: An experimental-phonetic approach*. Cambridge: Cambridge University Press.

Trubetzkoy, N. S. (1958). *Grundzüge der Phonologie* [Principles of Phonology]. Göttingen: Vandenhoeck und Ruprecht.

Truckenbrodt, H. (2004). Final lowering in non-final position. *Journal of Phonetics*, *32*, 313–348. doi:10.1016/j.wocn.2003.11.001

Tseng, C. Y., Pin, S. H., Lee, Y. H., Wang, H. M., & Chen, Y. C. (2005). Fluent speech prosody: Framework and modeling. *Speech Communication*, *46*, 284–309. doi:10.1016/j.specom.2005.03.015

Tseng, C.-Y., Su, Z.-Y., & Lee, L.-S. (2009). Mandarin spontaneous narrative planning—prosodic evidence from national Taiwan university lecture corpus. In *Proceedings of Interspeech* (pp. 2943–2946). Brighton.

Turk, A. E., & Shattuck-Hufnagel, S. (2007). Multiple targets of phrase-final lengthening in American English words. *Journal of Phonetics*, *35*, 445–472. doi:10.1016/j.wocn.2006.12.001

van Zyl, M., & Hanekom, J. J. (2012). When "okay" is not okay: Acoustic characteristics of single-word prosody conveying reluctance. *Journal of the Acoustical Society of America*, *133*(1), EL13–EL19. doi:10.1121/1.4769399

Varga, L. (1998). Rhythmical variation in Hungarian. *Phonology*, *15*, 227–266. doi:10.1017/S0952675798003583

Volskaya, N., & Stepanova, S. (2004). On the temporal component of intonational phrasing. *Proceedings of Speech and Computer (SPECOM)*, 641–644. St. Petersburg, Russia, St. Petersburg Institute for Informatics and Automation of RAS.

Vroomen, J., Tuomainen, J., & de Gelder, B. (1998). The roles of word stress and vowel harmony in speech segmentation. *Journal of Memory and Language*, *38*, 133–149. doi:10.1006/jmla.1997.2548

Wagner, M. (2005). *Prosody and recursion* (Unpublished doctoral dissertation). Cambridge, MA: Massachusetts Institute of Technology.

Wagner, M. (2010). Prosody and recursion in coordinate structures and beyond. *Natural Language and Linguistic Theory*, *28*, 183–237. doi:10.1007/s11049-009-9086-0

Wagner, M., & Watson, D. (2010). Experimental and theoretical advances in prosody: A review. *Language and Cognitive Processes*, *25*, 7–9, 905–945. doi:10.1080/01690961003589492

Wang, B., & Xu, Y. (2011). Differential prosodic encoding of topic and focus in sentence-initial position in Mandarin Chinese. *Journal of Phonetics*, *39*, 595–611. doi:10.1016/j.wocn.2011.03.006

Ward, G., & Hirschberg, J. (1985). Implicating uncertainty: The pragmatics of fall-rise intonation. *Language*, *61*, 747–776. doi:10.2307/414489

Warren, P., & Daly, N. (2000). Sex as a factor in rises in New Zealand English. In J. Holmes (Ed.), *Gendered speech in social context: Perspectives from town and gown* (pp. 99–115). Wellington: Victoria University Press.

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

Watson, D. G., Tanenhaus, M. K., & Gunlogson, C. (2008). Interpreting pitch accents in on-line comprehension: H* vs. L_H*. *Cognitive Science*, *32*, 1232–1244. doi:10.1080/03640210802138755

Weber, A., Braun, B., & Crocker, M. W. (2006). Finding referents in time: Eye-tracking evidence for the role of contrastive accents. *Language and Speech*, *49*, 367–392. doi:10.1177/00238309060490030301

Weber, F., Manganaro, L., Peskin, B., & Shriberg, E. (2002). Using prosodic and lexical information for speaker identification. *Proceedings of International Conference on Acoustics, Speech, and Signal (ICASSP)*, *1*, 141–144.

Wennerstrom, A. (2001). *The music of everyday speech*. Oxford: Oxford University Press.

Wichmann, A. (2011). Prosody and pragmatic effects. In G. Anderson & K. Aijmer (Eds.), *Pragmatics in society* (181–213). Berlin: DeGruyter Mouton.

Wichmann, A., House, J., & Rietveld, T. (2000). Discourse effects on f0 peak alignment in English. In A. Botinis (Ed.), *Intonation: Analysis, modelling and technology* (pp. 163–184). Dordrecht: Kluwer Academic.

Wightman, C. W., Shattuck-Hufnagel, S., Ostendorf, M., & Price, P. J. (1992). Segmental durations in the vicinity of prosodic phrase boundaries. *Journal of the Acoustical Society of America*, *91*, 1707–1717. doi:10.1121/1.402450

Xu, Y. (1999). Effects of tone and focus on the formation and alignment of F0 contours. *Journal of Phonetics*, *27*, 55–105. doi:10.1006/jpho.1999.0086

Xu, Y. (2010). In defense of lab speech. *Journal of Phonetics*, *38*, 329–336. doi:10.1016/j.wocn.2010.04.003

Xu, Y., & Wang, E. (2001). Pitch targets and their realization: Evidence from Mandarin Chinese. *Speech Communication*, *33*, 319–337. doi:10.1016/S0167-6393(00)00063-7

Xu, Y., & Xu, C. X. (2005). Phonetic realization of focus in English declarative intonation. *Journal of Phonetics*, *33*, 159–197. doi:10.1016/j.wocn.2004.11.001

Yoon, T.-J. 2007. *A predictive model of prosody through grammatical interface: A computational approach* (Unpublished doctoral dissertation). University of Illinois, Urbana-Champaign.

Yoon, T., Chavarría, S., Cole, J., & Hasegawa-Johnson, M. 2004. Intertranscriber reliability of prosodic labeling on telephone conversation using ToBI. In *Proceedings of Interspeech 2004* (pp. 2729–2732). Jeju.

Yoon, T., Cole, J., & Hasegawa-Johnson, M. (2007). On the edge: Acoustic cues to layered prosodic domains. In *Proceedings of the 16th International Congress of Phonetics Sciences* (pp. 1264–1267). Saarbrücken.

Yule, G. (1980). Speakers topics and major paratones. *Lingua*, *52*, 3–47. doi:10.1016/0024-3841(80)90016-9