White, L. (2018). Segmentation of speech. In S.-A. Rueschemeyer & G. Gaskell (Eds.), *Oxford Handbook of Psycholinguistics* (pp. 5-30). Oxford: Oxford University Press.

Chapter 1: Segmentation of Speech

Laurence White

1 Introduction

The power of words may have been overstated in communication research. A useful, reductivist schema for spoken interaction is a linear information chain: concepts in the speaker's mind are encoded into a stream of physical sounds, picked up by the ears of listeners and appropriately decoded in their minds. This is closely analogous to Morse code telegraphy: an active, encoding transmitter and a passive, decoding receiver. Although Morse code is now dustily antiquated as a communication system, the information transmission model has been implicit in much – highly productive – psycholinguistic research.

Morse code is famously a binary digital system, but the sequence of dots and dashes that represent letters – *dot dot dot dash dash dash dot dot dot* for "SOS" – are delimited by silent intervals, short between letters and longer between words. Morse signals would not be interpretable without this explicit segmentation. These aspects of electrical telegraphy have informed assumptions about speech comprehension: firstly, that identification – decoding – of the individual words of a message is the primary goal of spoken interaction; secondly, that locating boundaries between words is a prerequisite to lexical decoding.

Spoken communication entails much more than physical encoding and perceptual decoding of word sequences, of course. Speech is typically produced in an interactive, reciprocal and negotiated context, with speaker and listener roles shifting in a dynamic process only partially mediated by sentential meaning. Prosody – particularly modulation of pitch and timing – conveys information about not only linguistic meaning but also turn-taking, attitudes, and physical and emotional states. Critical information also flows through channels physically distinct from the acoustic stream, notably visual, but potentially also haptic and proprioceptive.

Words, therefore, are not all there is to speech. And yet words themselves are not illusory. Early infant forays into language production demonstrate that words are true building blocks of communication, and a full account of speech processing must make reference to how we extract words and associate them with meaning. The first stage of this process, speech segmentation, is often characterized by reference to the absence of consistent boundaries between spoken words, by

contrast with the white spaces between words on a page or screen of text. Furthermore, speech sounds themselves are contextually variable. In Sections 2 and 3, we review how research on the problems of continuity and variation has reframed concepts of speech representation and informed the search for universal segmentation mechanisms against a background of linguistic diversity.

The continuity problem and the information transmission model imply that explicitly locating word boundaries is a prerequisite for speech understanding. Given clear, unambiguous input, however, lexical knowledge can provide segmentation implicitly, as discussed in Section 4. In the face of ambiguity or signal degradation, listeners exploit a range of "non-lexical" (segmental and prosodic) segmentation cues, reviewed in Section 55. Interactions between implicit, lexically-driven segmentation and explicit use of non-lexical cues are discussed in Section 6. We note that much of the experimental work reviewed here pertains to languages of European origin, a reflection of the extant literature in English: whilst specific non-lexical cues may have more relevance for some languages than others – lexical stress or vowel harmony, for example – the primary aim is to show the language-general mechanisms whereby different sources of segmentation information are weighted and integrated by listeners.

The requirement for controlled speech stimuli in laboratory perception studies neglects the dynamic nature of speaker-listener interactions, but in Section 7 we consider the active role of infant-directed speech in facilitating young learners' extract of words from the speech stream, and briefly examine the parallels and contrasts between first language and second language segmentation.

2 Problems in speech segmentation: Continuity and variation

We experience the continuity of speech when overhearing fluent conversation in an unfamiliar language. Whilst intonation and gesture offer clues to the nature of the interaction – particularly its social and emotional significance – individual words are rarely obvious, except when highlighted by rhetorical devices such as repetition or exaggerated pausing. The momentary recognition arising from sudden code-switching into a familiar language (e.g., "*Je vais parler aux builders ce soir*") strikingly emphasizes the power of lexical knowledge to impose discrete structure on continuity.

Even in familiar languages, however, lexical access cannot simply be a matter of mapping sound sequences to stored representations (Norris & Cutler, 1985). The scarcity of unambiguous boundary markers means that new lexical candidates could potentially be activated with every new segment, causing a combinatorial explosion in word hypotheses. The variable realization of speech sounds

also introduces uncertainty: indeed, unless we can generalize from phonetic variability, connecting two repetitions of a word with their common underlying representation would be highly problematic.

An abstract level of representation, intervening between acoustic processing and the search for lexical matches, therefore has intuitive appeal. Norris and Cutler (1985) draw a theoretical contrast between "classification" and "segmentation". Classification entails dividing speech into a series of labelled units, of a particular type, to produce a phonological specification, e.g., a string of phonemes or syllables. Classification must be exhaustive – no parts of an utterance should be left unclassified – and there should be some systematic mapping from phonological units to words themselves. One theoretical attraction of an intervening, sublexical level of representation is its power to constrain the search space: any spoken language could encode tens of thousands of words, but a few dozen phonemes and a few hundred different syllables. Indeed, segmenting continuous speech into discrete words is potentially boosted by sublexical classification, particularly into a sequence of syllables. Syllable boundaries represent points of look-up for new lexical candidates, but that process need not be initiated within syllables. As discussed below, however, claims about the power and ubiquity of the syllable as a perceptual unit have been empirically challenged.

3 Problems in speech segmentation: The syllable in prelexical classification

Much early experimental psycholinguistics focused on parsing of the speech stream into units of perception, "building blocks of prelexical processing… [generating] a transcript of the signal as a sequence of units (e.g., phonemes, demi-syllables, syllables)" (Mattys & Melhorn, 2005, p. 224). In a key study, Mehler, Dommergues, Frauenfelder and Segui (1981) used word-monitoring latencies to assess French listeners' sensitivity to syllable structure. After visual presentation of the target *pa*, for example, participants were quicker to detect it in *palace* than in *palmier*, but quicker to detect *pal* in *palmier* than in *palace*. This crossover interaction was interpreted as facilitation of detection when the target (*pa* vs *pal*) exactly corresponded to a syllable, based on French syllabification *pa.lace* vs *pal.mier*. Mehler et al. thus proposed that the syllable constitutes a perceptual unit of processing; furthermore that syllabification is prior to, and necessary for, lexical access.

The "syllable effect" interpretation was challenged by similar studies with English listeners. English has a larger range of syllable structures than French, with multiple consonants allowed in onset and coda clusters, and appears less straightforward to syllabify. For example, the medial consonant in strong-weak words like *balance* is typically regarded as ambisyllabic (see Table 1 for definitions of phonetic terms), evidenced by native English speakers' variable intuitions about *ba.lance* or *bal.ance*

as the appropriate syllabification (Anderson & Jones, 1974; Cutler, Mehler, Norris, & Segui, 1986). Indeed, English listeners did not demonstrate the syllable effect, but showed – for both open and closed targets, e.g., *ba* and *bal* respectively – quicker detection in words with ambisyllabic medial consonants (e.g., *ba[l]ance*) than unambiguous syllable structures (e.g., *bal.cony*) (Cutler et al., 1986).

Interpreting different French and English responses to equivalent materials, Cutler et al. (1986, p. 397) argued that: "language-specific components [of a psycholinguistic theory] are highly undesirable; if language specificity at this level is possible, why not dialect specificity or even speaker specificity? […] For this reason we feel compelled to suggest a language-universal framework [...]" Specifically, they proposed that listeners may have both phonemic and syllabic segmentation strategies available, mediated by experience and task demands. Thus, their English results suggest that simple CVCV structures may boost prelexical phonemic classification, but this is overridden when unambiguous native language syllabification experience biases listeners' decisions, as in Mehler et al. (1981) for French.

Cutler et al. (1986) ruled out a role for lexical stress in biasing English listeners' responses, because they showed the same bias – quicker for both *ba* and *bal* in *balance* than *balcon* – in French words with consistent final accent (contrary to predominant English stress, Section **Error! Reference source not found.**). Stress came to the fore subsequently, however. Cutler and Norris (1988) found that English listeners' detection of, e.g., *mint* in strong-weak (SW) *mintef* was faster than in strong-strong (SS) *mintayf*, but *thin* was no more quickly detected in *thintef* than *thintayf*. They concluded that strong syllables – specifically, full vowels – trigger a search for new lexical items beginning with a matching syllable (see Table 1 for working definitions of phonetic terms). Because medial /t/ in SS contexts syllabifies as a syllable onset (-*tavye*), it activates compatible lexical representations and interferes with activation of *mint*, for which the same /t/ is required. Weak syllables would not initiate lexical access: thus the representation of the /t/ as word-final in *mint* is not blocked by *mintef*.

When the target is wholly contained within the first syllable – *thin* in both *thintayf* and *thintef* – initiation of lexical access by the following syllable becomes irrelevant (Cutler & Norris, 1988). The precise matching of the target with a complete syllable – the first syllable in the context being *thin* or *thint* – does not matter without a competing attempt at lexical access. This represents a theoretical shift from Mehler et al. (1981). Cutler and Norris see no need to predicate lexical access on

ambisyllabic	Ambisyllabic consonants are regarded as belonging to both the preceding and following syllable, as the /l/ in English *balance*.
full vowel / *reduced vowel*	Full vowels are produced with a non-centralized articulation and are typically of significantly longer duration than reduced/centralized vowels. In English, full vowel include monophthongs (e.g., the vowels in *seat*, *set*, *sat*) and diphthongs (e.g.,the vowels in *boat*, *bait*, *bout*). Reduced vowels include schwa, as in the first syllable of *commend*, and [I] in unstressed syllables like the second syllable of *captain*.
hyperarticulation/ *hypoarticulation*	Hyperarticulation is the exaggeration of phonetic gestures so as to produce speech sounds that are maximally distinctive, as in clear speech styles. Hypoarticulation is the reduction, assimilation or elimination of sounds, as in casual speech styles, through which some phonemic distinctions may be lost.
lexical stress	In phonetic terms, lexical stress is prominence of a particular syllable within a word, conveyed through lengthening, loudness and pitch.
phrasal accent	In phonetic terms, phrasal accent is prominence of a particular syllable or word within a phrase, conveyed through pitch, lengthening and loudness.
strong syllable / *weak syllable*	In English, a strong syllable contains a full vowel, as in the first syllable of *captain* or the second syllable of *commend*; a weak syllable contains a reduced vowel or no vowel, as in the first syllable of *commend* or the second syllables of *captain* and *bottle*. Most strong syllables in English are lexically stressed, but some are unstressed, e.g., the second syllable in *insight*.

Table 1: Glossary of some phonetic terms used in this review. These are intended as heuristics for understanding rather than theoretical definitions.

exhaustive classification into sublexical units: rather, certain events within utterances trigger the initiation of hypothesis-testing about words (Cutler, McQueen, Norris, & Somejuan, 2001) and languages differ in their provision of such triggering events. The *universal* requirement is that word

boundaries are located; the language-specific variation is not in the units of sublexical processing, but rather the strategies available for triggering boundary hypotheses.

Content, Meunier, Kearns and Frauenfelder (2001) presented a further challenge to the syllable effect, noting that Sebastián-Gallés, Dupoux, Segui and Mehler (1992) only found the effect for Catalan speakers with unstressed syllables, and found no effect at all for Spanish speakers unless response times were slowed with a parallel semantic task. In non-slowed trials, Spanish speakers showed an overall faster response for CV targets over CVC targets (also found for Italian listeners, Tabossi, Collina, Mazzetti, & Zoppello, 2000). Content, Meunier, et al. themselves found a syllable effect in French only where the pivotal consonant was a liquid, e.g., *ba[l]ance*, *ba[l]con* – as in Mehler et al., (1981) – but even for such stimuli the effect relied on specific test-trial blocking and relative response slowness. Like Cutler and Norris (1988), they rejected a mandatory prelexical syllabification stage, interpreting results in terms of phoneme-level matching, informed by allophonic and coarticulatory information (Content, Meunier, et al.). Significantly, they propose that incoming speech information is used for lexical access as it becomes available.

Content, Kearns and Frauenfelder (2001) break altogether with the conception of segmentation as a process of identifying boundaries between adjacent constituents. Rather, they emphasize that identification of potential lexical onsets is prior to, and distinct from, judgements about offsets. Theoretically, the hierarchical, recursive nature of linguistic structure means that a constituent can begin before another constituent of the same type has ended. Empirically, Content, Kearns, et al., cite metalinguistic syllabification tasks, where French participants' judgements about syllable codas, but not syllable onsets, were influenced by stimulus characteristics and task demands.

Regarding segmentation, the primary proposal of Content, Kearns, et al., (2001) is that the onset of any strong (full vowel) syllable may be taken to be a possible word onset. (The strong syllable qualification accounts for differences between French and English studies, given that French does not manifest widespread vowel reduction.) This proposal presupposes that listeners are sensitive to syllabic structure, but does not require an exhaustive classification into contiguous, non-overlapping syllables. Accordingly, Dumay, Frauenfelder and Content (2002) found that misalignment of target syllable onsets in nonsense carrier strings was more problematic for word-spotting than misalignment of syllable codas: e.g., participants were slower to spot *lac* in misaligned /zy.glak/ than /zyn.lak/, but no difference between /lak.tyf/ and misaligned /la.klyf/. Similar asymmetries between

onset and coda misalignment were demonstrated for Dutch (McQueen, 1998; Vroomen & De Gelder, 1997).

The Syllable Onset Segmentation Heuristic (SOSH) captures a theoretical shift: the critical information for lexical access is the location of possible word onsets (Content, Kearns, et al., 2001). Which syllable onsets activate possible words is proposed to be mediated by language-specific factors such as metrical structure. Dumay et al. (2002, p. 14) further observe that "syllable-based segmentation strategies such as SOSH […] are not deterministic rules but heuristics, so that their effect could be modulated or compensated by other cues, such as lexical information." As discussed in Section 6, where lexical information about structure is adequately informative, it may render non-lexical heuristics superfluous.

4 Segmentation strategies: Word recognition and implicit segmentation

4.1 Models of word recognition

Models of word recognition are typically based on simultaneous activation of multiple lexical candidates and competition between those candidates: e.g., TRACE (McClelland & Elman, 1986); Shortlist (Norris, 1994; Norris & McQueen, 2008). Models often have multilevel connectionist architecture, with discrete nodes at one or more levels of sublexical representation (e.g., features and/or phonemes) together with nodes corresponding to individual words. Phonetic input activates sublexical representations which then excite compatible lexical nodes, whilst incompatible lexical nodes are inhibited. Levels of activation are determined both by bottom-up goodness of fit and competition – via inhibitory connections – at the word level (models such as TRACE also allow top-down excitation, McClelland & Elman, 1986). Eventually, competition results in recognition of the most strongly activated candidate word. Given the temporal nature of speech, the whole process is continuous and overlapping, with further word candidates being activated as new phonetic input is received.

Segmentation emerges naturally from the outcome of lexical competition. For example, the phrase "*silver doorbell"* provides phoneme-level evidence to temporarily activate the boundary-straddling *adore*. As the activation of the optimally-matching sequence ultimately prevails, the boundary ("…*ver#door*…") is respected, but this veridical segmentation is achieved without explicit reference to boundary-specific features of the signal.

Models differ, however, in whether segmentation cues explicitly inform lexical hypotheses. TRACE identifies words purely through activation between levels and inhibition within levels (McClelland & Elman, 1986). By contrast, Shortlist allows certain boundary-relevant constraints to influence activation: in particular, the Possible Word Constraint (PWC) requires that lexical solutions do not leave stray segments – specifically vowel-free syllables – that could not be viable words (Norris, McQueen, Cutler, & Butterfield, 1997). Thus, in "*red frock*", activation of the sequence "*red rock*" is disfavoured because this strands a lone /f/, not a possible English word. Metrical information also constrains activation in some instantiations of Shortlist: Norris, McQueen and Cutler (1995) proposed a segmentation boost to all (English) lexical candidates with onsets aligned to strong syllables (excepting that all post-silence syllables must be word-initial). Subsequently, Norris et al. (1997) integrated this metrical constraint with the PWC, penalising lexical candidates that leave impossible words (i.e., stray consonants) between hypothesized word edges and boundaries in the signal (silences and strong syllable onsets): thus *apple* is harder to detect in *fapple* than in *vuffapple*.

Whether, and how much, explicit segmentation information is utilized in word recognition is debatable. The time-course of lexical activation indicates that candidates disfavoured by non-lexical cues (timing, allophony, phonotactics, etc.) are still considered. Embedded words – e.g., subset *bone* in superset *trombone* – are activated (Shillcock, 1990), with activation modulated by degree of subset-superset overlap (Bowers, Davis, Mattys, Damian, & Hanley, 2009). Boundary-straddling words are also activated, e.g., Italian *visite* ("*visits*") in response to *visi tediati* ("*bored faces*"), even when strong distributional cues disfavour the overlapping word (Tabossi et al., 2000; see also Gow & Gordon, 1995). Whilst overlapping words are relatively rare, embedded words abound: McQueen, Cutler, Briscoe and Norris (1995) reported that 84% of English polysyllables contain at least one shorter word, most often with coincident onsets (e.g., *mace* in *masonry*), although – as in *bone*/*trombone* – later embedded words are also activated.

Thus, non-lexical segmentation cues do not serve to rule out alternative parses of the input signal at an early stage, but lexical competition has a strong effect. Target words embedded in contexts with no competitor words are recognized more quickly, e.g., *sack* is detected more quickly in /sAkr@k/, with no competitors, than /sAkr@f/, potentially completed as *sacrifice* (McQueen, Norris, & Cutler, 1994). Moreover, graded effects of the number of competitors are also found: specifically, detection of targets is impeded with increasing cohort size of alternative parses (Vroomen & de Gelder, 1995, see Section 5.2).

By contrast, non-lexical cues to segmentation are strengthened by the cumulative weight of lexical evidence, the theoretical power of English metrical segmentation being just one example (Cutler & Carter, 1987). Non-lexical segmentation cues should be more reliable given greater numbers of words with equivalent features, in contrast to the negative effects of neighbourhood on word recognition. Explicit segmentation and word recognition thus appear distinct, influenced by contrasting statistical contingencies: "Information on what is a word in the language appears to behave quite differently from information on what is likely to be a word in the language" (Newman, Sawusch, & Wunnenberg, 2011, p. 474).

4.2 Evidence for the role of lexicality, syntax and semantics in segmentation

There is abundant support for the power of lexical knowledge to impose structure on speech. For example, listeners' ability to extract new words from artificial language streams is enhanced when the stream includes already familiarized nonwords (Cunillera, Càmara, Laine, & Rodríguez-Fornells, 2010; Dahan & Brent, 1999). This "segmentation-by-lexical-subtraction" strategy accords with evidence from cross-modal fragment priming (Mattys, White, & Melhorn, 2005): in this paradigm, a lexical decision (respond "word" or "nonword") to a visually-presented trisyllabic target (e.g., "*corridor*") is preceded by a five-syllable auditory stream containing a trisyllabic context and the first two syllables of the target (e.g., *anythingcorri*). Mattys et al. found that where the auditory context was a word (e.g., *anything[corri]* vs nonword *imoshing[corri]*), there was faster lexical decision to the target, indicating more effective segmentation of the auditory prime *corri*. Thus, the lexical status of a context word promotes the extraction of the subsequent – contiguous, non-overlapping – word, in accordance with the concept of implicit segmentation through word recognition. Using the same paradigm, Mattys et al. further found that, where lexicality provided a segmentation solution, a range of non-lexical cues – stress, phonotactics, decoarticulation – were neglected by listeners (Section 6).

Despite the power of segmentation-by-lexical subtraction, Vroomen and de Gelder (1995) found no effect of the *wordlikeness* of nonwords that were contiguous with, but did not overlap, the target: e.g., target BEL in *belkem* vs *belkeum* vs *belkaam* (in increasing order of cohort size of the second syllable, *k_m*). Likewise, (Newman et al., 2011) found no effect of lexical neighbourhood of the preceding nonword syllable in a word-spotting task. Identifying a word in an utterance creates edges in the speech stream, but the cohort effects that are a feature of recognition – specifically, the plausibility of nonwords – do not influence boundary perception where nonwords do not overlap with lexical candidates.

Cunillera, Laine and Rodríguez-Fornells (2016) demonstrated a neurophysiological correlate of the power of lexical knowledge: using artificial language learning, they found that familiarized words in the language stream elicited greater stimulus-preceding negativity (SPN), a frontal event-related potential (ERP) associated with expectation for relevant upcoming information. Cunillera et al. infer that this activity is an index of orientation to the subsequent novel word in the stream, also finding that SRN magnitude decreases with better recognition of the novel words: thus, upcoming segmental material elicits lower activity as the language becomes more familiar.

Listeners' interpretations are, of course, influenced by syntactic structure, semantics, pragmatics and contextual factors beyond the signal (Cole, Jakimik, & Cooper, 1980). Although the acoustic evidence may permit ambiguity, on hearing "*Time and tide wait for…*", we are unlikely to conjure the mysterious character "*Gnome Ann*" from the subsequent segments. Likewise, "*four candles*" is a pragmatically more plausible hardware shop request than "*fork [h]andles*". Reviewing the evidence, however, McQueen (2005) suggests that foregoing context does not influence what lexical representations are initially activated, but rather the outcome of lexical competition. This secondary role for context accords with Mattys, Melhorn, & White (2007), who modulated acoustic and syntactic evidence for the parsing of ambiguous phrases such as *takes pins* vs *take spins*. The latter interpretation, when given a foregoing plural subject (e.g., *Those women take spins*), was strongly favoured whatever the acoustics, showing a clear role for context. The singular context (e.g., *That woman takes pins*) was much less constraining, however, and acoustics dominated the parse. Critically, in the plural case, the lexical affiliation of the ambiguous /s/ was late and so context could influence activation, but where /s/ was affiliated to the preboundary verb, the acoustics were already adjudicating on the ambiguity.

Some investigations of the role of speakers in actively modulating explicit cues to segmentation actually reinforce the power of syntactic and segmental context over acoustics. Kim et al. (2012) found that speakers provided minimal acoustic cues to the segmentation of ambiguous schwa-initial sequences (e.g., *a door* vs *adore*), and those cues that were available did not guide listeners to discriminate the alternatives. Prior syntactic and semantic context favouring one or other reading (e.g., *The hallway leads to…*) were highly biasing, however.

Kim et al. (2012) used an offline metalinguistic judgement task, e.g., "Did you hear one word or two?" which could have biased listeners to favour context over acoustics. The power of higher-level

information was, however, reinforced in a cross-modal identity-priming study (White, Mattys, & Wiget, 2012b), examining how acoustic segmentation cues were modulated by speech style: specifically, spontaneous map task speech vs read versions of the same utterances. In two-word phrases with no juncture ambiguity, but contrasting semantic predictability (e.g., high predictability – *oil tanker*; low predictability – *seal tanker*), there was relative hypoarticulation of acoustic cues in the map descriptions compared to read speech. However, despite the contrasting acoustic cue strength, there was no effect of speech style on the cross-modal priming these phrases elicited, whilst semantic relatedness boosted priming (White et al., 2012b). As in Kim et al.'s study, available lexical and semantic cues dominated acoustics (see also Mattys et al., 2005, Experiment 5).

5 Segmentation strategies: The role of non-lexical cues

Speech production studies demonstrate that the signal provides diverse potential segmentation cues. Many cues are language-specific, including allophony conditioned by word and phrase boundaries, and distributional patterns, notably consonant phonotactics and vowel harmony. Such patterns must be inferred from linguistic experience and some degree of phonological generalization: vowel harmony-based segmentation in Finnish, for example, requires listeners to categorically distinguish front and back vowels (Suomi, McQueen, & Cutler, 1997). Prosodic cues require broader abstraction: to apply metrical segmentation, learners must categorize syllables as strong or weak, and learn their distributional regularities. Whether there are segmentation cues so general as to apply universally, regardless of language-specific experience, remains an open question. Phonetic and perceptual evidence suggests, however, that – apart from between-word pauses – prosodic edge cues, reviewed below, are the strongest candidates for universality.

Potential cues to word boundaries are not deterministic. Even silent pauses can occur within words in disfluent speech. To exploit non-lexical segmentation cues, therefore, listeners must not only generalize about the phonological categories to which they apply, but also infer heuristics about cue reliability. For example, most English content words begin with strong syllables, but strict metrical segmentation would cause frequent misparsing (Harrington, Watson, & Cooper, 1989). Furthermore, the strength of acoustic-phonetic or prosodic cues varies according to speaking context, speech rate and speaker awareness of ambiguity and listeners' needs (Lindblom, 1990). And, despite the potential of audience design considerations to account for variation, speakers do not always provide disambiguating information (Kim et al., 2012), nor do listeners necessarily exploit all available cues (White et al., 2012b).

Production and perception studies of non-lexical segmentation cues are reviewed below, focusing first on those that may pertain regardless of listeners' native language(s).

5.1 Language-general cues

Pauses

It is a commonplace that boundaries between words in fluent speech are rarely associated with pauses, although we perceive familiar languages as discrete words. One function of infant-directed speech may be to break up utterances into short or even single-word prosodic phrases to promote segmentation (Section 7.17.1). Adult-directed speech is also characterized by shorter, pause-delimited phrases in hyperarticulated clear speech style, with concomitant intelligibility benefits (Smiljanić & Bradlow, 2008), although speakers vary in clear speech strategies and how far they assist the listener (Smiljanić & Bradlow, 2009).

Where available, even subliminally short between-word pauses promote learning in artificial languages (Peña, Bonatti, Nespor, & Mehler, 2002), and longer, perceptible pauses are associated with ERP signatures of word learning (Mueller, Bahlmann, & Friederici, 2008). Thus, as expected, pauses promote segmentation in novel linguistic conditions (also Finn & Hudson Kam, 2008), and clear speech styles associated with difficult listening conditions manifest greater pause frequency. However, constraints on communicative efficiency, the power of implicit segmentation through lexical recognition and the availability of other non-lexical cues all serve to limit pause frequency in typical conversational contexts.

Cross-boundary decoarticulation

The speech continuity problem arises not only because words are contiguous, but because sounds are coarticulated: the phonetic realization of phonemes depends upon preceding and following segmental material (Öhman, 1966). Articulatory gestures are, however, strengthened immediately before and after word boundaries, entailing less gestural overlap, and this decoarticulation increases with prosodic boundary strength (Fougeron & Keating, 1997). Such articulatory strengthening is also associated with lengthening; for word-initial consonants, both are widely observed across languages (Keating, Cho, Fougeron, & Hsu, 2004).

Decoarticulation is interpreted as a boundary cue in artificial language learning, carrying more weight in clear speech than syllable transition probabilities (Fernandes, Ventura, & Kolinsky, 2007); in noisy speech, however, transition probabilities dominate, presumably because subtle acoustic

variations associated with decoarticulated boundaries were masked. In a cross-modal fragment priming paradigm, Mattys (2004) showed the power of boundary decoarticulation over lexical stress cues in English, again with the pattern reversed in noise.

Prosodic lengthening

Edges of words and higher-level prosodic domains are associated with segmental lengthening. Onset consonants are longer word-initially than medially (e.g., Oller, 1973) and greater lengthening is observed phrase-initially (Fougeron & Keating, 1997; although utterance-initial consonants are often acoustically short, see White, 2014, for a functional interpretation). Pre-boundary lengthening of vowels and coda consonants is widely observed across languages (e.g., Berkovits, 1991; Wightman, Shattuck-Hufnagel, Ostendorf, & Price, 1992). Whether word-final lengthening is observed phrase-medially is debatable; certainly, any preboundary lengthening is attenuated in the absence of a phrase boundary (White & Turk, 2010). With regard to the heads of prosodic domains: in addition to lengthening of lexically-stressed syllables, both stressed and unstressed syllables are lengthened within phrasally-accented words (Turk & White, 1999). Accentual lengthening of the primary stressed syllable is attenuated in longer words, giving rise to observations of so-called "polysyllabic shortening" – e.g., *cap* longest as a monosyllable, shorter in *captain*, shorter still in *captaincy* – an effect which is minimal in the absence of phrase-level lengthening effects (White & Turk, 2010).

Preboundary lengthening, particularly of vowels, clearly promotes boundary detection (e.g., Price, Ostendorf, Shattuck-Hufnagel, & Fong, 1991; Saffran, Newport, & Aslin, 1996); indeed, this is claimed to be a universal cue (Tyler & Cutler, 2009; but see Ordin, Polyanskaya, Laka, & Nespor, 2017). Lengthening of word-initial consonants boosts English listeners' segmentation of artificial languages (White, Mattys, Stefansdottir, & Jones, 2015; see Shatzman & McQueen, 2006, regarding initial consonant lengthening and Dutch segmentation). Moreover, the locus of lengthening is important: where the vowel in the initial syllable is lengthened rather than the consonant, segmentation is not boosted (White et al., 2015). Shortening of stressed syllables in longer words (due to attenuation of prosodic lengthening, White & Turk, 2010) helps rule out lexical embedding (e.g., *ham* vs *hamster*; Davis, Marslen-Wilson, & Gaskell, 2002; Salverda, Dahan, & McQueen, 2003). Finally, faster foregoing speech rate increases the power of boundary lengthening cues, suggesting that listeners dynamically adjust predictions about the timing of speech events (Reinisch, Jesse, & McQueen, 2011).

Cross-boundary glottalization

When words lack vocalic onsets, the parallel of initial lengthening/strengthening is initial vowel glottalization (also known as creaky voice/laryngealization/vocal fry), realized as irregular glottal pulses, often with amplitude reduction (Dilley, Shattuck-Hufnagel, & Ostendorf, 1996). Many languages manifest glottalization at vowel-to-vowel word boundaries; furthermore, the incidence of pre- and post-boundary glottalization increases with boundary strength (Dilley et al., 1996; Pompino-Marschall & Żygis, 2010). Listeners duly interpreted vowel glottalization as a cue to preceding boundaries (Nakatani & O'Connor-Dukes, 1980; Newman et al., 2011). In contrast with some other boundary-related allophonic variations, the qualitative nature of initial-vowel glottalization may make it particularly salient.

Intonational boundaries

Intonational patterns are highly variable cross-linguistically, but association of the heads and edges of prosodic structure with accents and boundary tones is very widely observed (Ladd, 2008). Thus, whilst intonational marking of phrase boundaries is language and context specific, the power of pitch events to cue phrasal segmentation is plausibly universal (e.g., Spinelli, Grimault, Meunier, & Welby, 2010, for French). Furthermore, certain intonational features – e.g., initial rise, falling pitch towards the final boundary – may be universally interpretable (Bolinger, 1964). Indeed, Shukla, Nespor and Mehler (2007) demonstrated that intonational boundary cues, combined with timing effects, could either reinforce or override statistical cues in artificial language learning, even when the intonational contours were non-native (Italian speakers listening to Japanese intonation).

5.2 Language-specific cues

Phonotactics

Languages are diversely restricted in how they construct syllables. Certain consonant sequences are permissible within syllable onsets and codas (e.g., English /st/, /sp/), whilst others are restricted to one syllable position: e.g., English onsets, but not codas, allow /tr/ or /fr/; codas but not onsets allow /nd/ and /mp/. Languages differ in phonotactic constraints: thus, /zm/ and /zb/ are well-formed onsets in Italian but not English.

Listeners are sensitive to native language phonotactics. Functional near-infrared spectroscopy showed more strongly left-lateralized fronto-temporal responses to German phonotactically legal nonwords (e.g., *brop*) than illegal nonwords (e.g., *bzop*); furthermore, ERP data indicated a stronger N400 to legal nonwords, suggesting that illegal words are discarded prelexically (Rossi et al., 2011).

Phonotactics constraints restrict possible syllabifications: e.g., the syllable boundary in *dovecote* must be between /v/ and /k/, as the sequence /vk/ cannot be an onset or coda. Using word-spotting, McQueen (1998) showed that phonotactics constraints influenced Dutch listeners' ease of detecting monosyllables: e.g., *pil* detected more quickly in *pil.vrem* than *pilm.rem* (periods indicating phonotactically legal syllabifications). The phonotactic effect was stronger for word-final targets: *rok* was spotted more quickly in *fiemrok* than *fiedrok*, as in Dutch /mr/ is not a legal onset or coda, whilst /dr/ must be an onset because /d/ does not occur as a coda. This asymmetry reinforces the importance of onsets for lexical access (Content, Kearns, et al., 2001).

(Newman et al., 2011) found that English constraints on word-final vowels did not influence segmentation: legal tense-vowel-final syllables and illegal lax-vowel-final syllables were equivalent in allowing following consonants to be realigned as codas (lax: *vuhf-apple*; tense: *veef-apple*). Skoruppa, Nevins, Gillard and Rosen (2015) reported, however, that the lax vowel constraint was used by English listeners segmenting nonsense words.

Knowledge of what sequences may occur within and between words suggests useful segmentation heuristics. However, converging lines of evidence indicate that listeners do not systemically exploit phonotactic knowledge where more direct sources of segmentation information are available. Harrington et al. (1989) showed that a segmentation algorithm based on trigram occurrence within vs between English words successfully detected only 37% of boundaries in a transcribed corpus. Furthermore, as observed by Mattys and Bortfeld (2015), phonemically-based distributional generalizations neglect the impact of connected speech processes, such as contextual allophony, assimilations, deletions and insertions.

Using cross-modal identity priming, White et al. (2012b) found that differences in distributional regularities of consonant diphones had no impact on segmentation. Two-word phrases – e.g., *cream rickshaw* vs *drab rickshaw* – contrasted in within-word and between-word frequencies of cross-boundary diphones: thus, /mr/ is vanishingly rare within English words, but not uncommon across boundaries; /br/ is frequent within words, but rare across boundaries. Even these strong contrasts in sequential frequencies had no impact on segmentation as indexed by cross-modal priming from the second word. White et al. proposed that phonotactic segmentation effects only appear in the absence of full lexical solutions. Similarly, Mattys et al. (2005) pitted lexicality and semantic

dependencies against phonotactic frequencies and found phonotactic effects emerged only when stimulus truncation delexicalized the materials.

The Possible Word Constraint (Section 4.14.1) was proposed as a language-universal mechanism (Norris et al., 1997). In particular, single consonants may not be left stranded in word recognition: e.g., the embedded word *right* is disfavoured in *shining bright* because the residual /b/ lacks another syllabic attachment. What is a minimal legal syllable varies between languages, however: Hanulíková, McQueen and Mitterer (2010) found no word-spotting penalty for stranding isolated consonants that are allowable words in Slovak.

Vowel harmony

Vowel harmony is a widespread phenomenon, notably in agglutinating languages. For example, the front/back feature of the first vowel in a Finnish word determines the allowable vowels throughout the word, which (unless one of two neutral vowels) must share the first vowel's frontness/backness. Finnish listeners can use vowel harmony for segmentation: for example, the word target *hymy* is detected more quickly in the disharmonious context *puhymy* than in the harmonious context *pyhymy* (Suomi et al., 1997). Exploitation of vowel harmony depends on linguistic experience: thus, Finnish, but not French or Dutch, listeners benefited from vowel harmony for learning an artificial language (Vroomen, Tuomainen, & de Gelder, 1998).

Allophony

Whilst domain-edge lengthening and initial vowel glottalization are widespread in the world's languages, other forms of positional allophony are language or dialect-specific. Newman et al. (2011) found that English consonant allophonic differences between word-initial and word-final position modulated segmentation behaviour according to consonant class: being pronounced as initial or final influenced segmentation more for voiceless stops and /l/ than for fricatives, which have weaker position-specific allophony. Positional timing effects (consonants longer initially than finally) tend to apply across consonant classes.

Lexical stress

Features of English stress converge to suggest a metrical segmentation strategy (Norris & Cutler, 1985). Firstly, consonants between strong and weak syllables may be ambisyllabic (e.g., *ho[ll]ow*, *fo[c]us*…), thus strong syllable onsets are more reliable boundary locations. Secondly, stressed syllables are more salient, and the segments therein more recognisable and informative about

lexical identity (Cutler & Foss, 1977; Huttenlocher & Zue, 1983). Thirdly, strong syllables tend to begin words in English. Cutler and Carter (1987) analysed the distribution of strong and weak syllables in the London-Lund corpus of English conversational speech. Cutler and Carter found that 90% of lexical (open-class) words were either monosyllables or polysyllables beginning with a strong syllable (i.e., containing a full vowel, the primary cue to English lexical stress, Fear, Cutler, & Butterfield, 1995). Furthermore, 74% of strong syllables were the initial or only syllables of lexical words, 11% were function (closed-class) word-initial, and 15% were non-initial in lexical or function words. Only 5% of weak syllables were lexical-word-initial. Accordingly, Cutler and Butterfield's (1992) analysis of juncture misperceptions found that English listeners were more likely to insert a spurious word boundary before a strong syllable than a weak syllable and – conversely – to delete a boundary before a weak syllable (hence errors such as "*Sheila Fishley*" for "*She'll officially*" and "*How bigoted?*" for "*How big is it?*").

In Dutch, like English, most content words begin with stressed syllables, thus a metrical segmentation heuristic is similarly plausible. Vroomen and de Gelder (1995) used cross-modal priming to investigate how stress and lexical competition affected segmentation. The materials were broadly analogous to the *mintef*/*mintayf* sets used by Cutler and Norris (1988), where activation of the target *mint* was blocked by the competing activation of the strong second syllable *tayf*, whilst weak second syllables do not trigger lexical access (Section 3). For example, Dutch listeners made lexical decision to visual targets (e.g., *melk*) heard at the offset of auditory contexts, in the critical trials, strong-weak (SW) *melkem*, SS *melkeum*, SS *melkaam*. Second-syllable cohort size was manipulated, with many more Dutch words beginning *kaa*- than *keu*- and cohort size negligible for the weak syllable in *melkem*. Cohort size was inversely related to the magnitude of priming effects, suggesting that lexical competition from the overlapping second syllable, rather than its stress status, is the critical factor.

Another series of cross-modal priming experiments (Mattys, 2004; Mattys et al., 2005), pitted metrical segmentation pairwise against other cues: decoarticulation, phonotactics, the lexical status of a context word preceding the target. In clear listening conditions, stress was ignored in favour of the other cues; however, against a background of noise, stress became effective, suggesting that stress is a fallback cue when more reliable sources of information are compromised. Indeed, native-language-congruent stress patterns – as cued by pitch accent – boost artificial language learning: Dutch and English, but not French, listeners benefitted from word-initial pitch cues (Tyler & Cutler, 2009), as did Dutch and Finnish, but not French listeners (Vroomen et al., 1998).

As noted above, vowel harmony was also available to segment the artificial languages of Vroomen et al. (1998): when words had initial stress, however, the impact of vowel harmony disappeared. This is contrary to findings for English – where stress is only relied when other cues are diminished (Mattys et al., 2005) – and suggests that in fixed-initial-stress language like Finnish, metrical segmentation may be weighted more highly. Cue weighting apparently correlates with consistency.

6 Segmentation strategies: Cues in combination

The accumulated evidence indicates the dominance of implicit segmentation through lexical recognition, suggesting that activation and competition provide a complete lexical solution wherever possible. Non-lexical segmentation cues are stochastic in nature: for example, the use of metrical segmentation in English would incur a significant error rate (Harrington et al., 1989). However, when required, listeners can exploit a range of non-lexical cues derived from prosodic and acoustic-segmental regularities with respect to boundaries. Thus, segmentation proceeds through dynamic, strategic exploitation of available information, with parsimonious processing of the potentially redundant encoding of boundaries in the signal: the cues that speakers provide are not necessarily what listeners use.

Mattys et al. (2005) proposed a hierarchical framework for segmentation, based on English data, identifying three tiers of information – from highest to lowest weighted – lexical, segmental and prosodic (i.e., metrical). Thus, word recognition in a familiar language can proceed wholly via the lexical tier and without reference to non-lexical segmentation cues at all. Non-lexical cues may be invoked due to ambiguity at the lexical level (as in cases of homophony: *grey tanker* vs *great anchor*) or, more generally, where words in an utterance are not yet represented in the listener's mental lexicon, as in first or second language acquisition. Alternatively, signal degradation and ambiguity due to articulatory imprecision or environmental noise may make adjudication between alternative lexical solutions impossible without recourse to explicit strategies. For English, Mattys et al. suggested that metrical segmentation is a last resort for native adult speakers, presumably because of its relative lack of reliability (Harrington et al., 1989).

Newman et al. (2011) suggested a refinement of the segmentation hierarchy, weighting cues within the segmental tier according to relative salience. Thus, they proposed that strong acoustic cues – word-initial vowel glottalization, word-initial consonant aspiration – carry more weight than distributional cues, such as syllable-final vowel phonotactics. This accords with suggestions that

phonotactic patterns are rarely exploited where other cues are available (White et al., 2012b). Newman et al. (2011) also suggested that exploitation of segmental cues may be modulated by the strength of speakers' production. The dynamic interaction between listener needs and speaker behaviour is explored below with respect to infant-directed speech.

7 Segmentation in first and second language acquisition

Although implicit segmentation often arises via word recognition for adults listening to their native language, infant first language (L1) and adult second language (L2) learners need other strategies through which to extract words and build vocabulary. For infants, this requirement has generated the hypothesis that acoustic events associated with word boundaries – in particular, the marking of the heads and edges of prosodic constituents – are critical in "bootstrapping" language acquisition (Fernald & McRoberts, 1996; Morgan & Demuth, 1996). Furthermore, the infant learner typically experiences an interactional style rather different from adult conversation. Evidence suggests that features of infant-directed speech (IDS) promote explicit segmentation and thus facilitate early word learning, with infants moving to implicit segmentation as vocabulary grows. The latter observation is also true of adult L2 learning, but there appears to be an interaction, for the mature learner, between non-lexical cues relevant to L1 and L2 segmentation.

7.1 The role of infant-directed speech

Dialogues between caregivers and infants demonstrate as clearly as any discourse context how speakers adjust their articulation to take account of perceived listener needs, in accordance with Lindblom's (1990) H&H hypothesis (hyperarticulation/hypoarticulation). Compared to typical adult speech, infants prefer infant-directed speech – slower rate; longer, more frequent pauses; substantial lengthening before prosodic boundaries; higher pitch; higher pitch range – particularly when these IDS features are exaggerated (Dunst, Gorman, & Hamby, 2012). Beyond simply engaging the infant's attention, IDS enhances segmental distinctiveness by expanding the vowel space, and highlights focused words through structure and prosody (e.g., Cristia, 2013). In addition, helping the child to extract individual words may be a critical functions of IDS (Thiessen, Hill, & Saffran, 2005).

IDS exposes infants to a greater number of isolated words than adult-directed speech (ADS), with learning of those words consequently enhanced (Brent & Siskind, 2001). Furthermore, IDS utterances are often realized as sequences of short phrases, providing at least one reliable edge for many words, with exaggerated final lengthening and boundary tones complementing or replacing silent pauses (e.g., Cristia, 2013). Indeed, infants distinguish aligned and misaligned syntactic and

prosodic phrasing in IDS, but not ADS (Nelson, Hirsh-Pasek, Jusczyk, & Cassidy, 1989), indicating sensitivity to boundary cues. Furthermore, suprasegmental cues to both boundaries and stress are exaggerated in IDS (Albin & Echols, 1996, for English; Fernald et al., 1989, for a range of languages). The relative weighting of suprasegmental cues varies cross-linguistically: in American English directed to 14-month-olds, Fisher and Tokura (1996) found utterance-internal phrase boundaries were associated with lengthening, whilst the primary phrasing cue in comparable speech of Japanese mothers was pitch variation.

Thus, infant-directed speech serves to boost young children's ability to extract words from speech, building their vocabulary and gradually facilitating a shift from predominantly explicit to adult-style, predominantly implicit, lexicon-driven segmentation.

7.2 Infant sensitivity to segmentation cues

Segmentation must precede word recognition in acquisition, but language-specific knowledge rapidly accumulates in typical development, and known words allow infants to infer the edges of flanking words (Brent, 1997). Bortfeld et al. (2005) demonstrated the importance of limited lexical knowledge for identifying boundaries in 6-month-olds, who recognized less familiar words when preceded by familiar words in infant-directed speech, e.g., *feet* in *Mommy's feet*.

For infants, however, lexically-driven segmentation must be complemented by non-lexical cues. In accordance with the metrical segmentation hypothesis, English-learning infants are sensitive to the distribution of stressed syllables: thus, 7.5-month-olds extract trochaic words (e.g., *kingdom*, *hamlet*) from utterances, but with iambic words (e.g., *guitar*) tend to recognize strong-weak boundary-straddling sequences (e.g., *taris,* from "*Your gui<u>tar is</u> in the studio*"; Jusczyk, Houston, & Newsome, 1999). English-speaking 8-month-olds also segment trochaic words from Italian utterances (Pelucchi, Hay, & Saffran, 2009): thus, despite prosodic differences, infants can apply metrical segmentation to extract phonotactically legal, but phonetically non-native words. This argues against the importance of "rhythm class" in determining segmentation strategies (Nespor, Shukla, & Mehler, 2011), as Italian and English have been proposed to belong to distinct classes (see White, Mattys, & Wiget, 2012a, for a review). Italian and English have contrastive lexical stress and predominant trochaic structure in common, however, and a series of studies of 8-month-old monolingual Canadians indicates the importance of familiarity with native metrical structure (Polka & Sundara, 2012). Canadian French-learners extracted iambic words (e.g., *beret*, *guitar*) from French, and Canadian English-learners extracted trochaic words (e.g., *hamlet, candle*) from English; however, both groups

failed to segment words from their non-native language. Here the importance of rhythm in early segmentation is to allow infants to induce the predominant metrical structure of their native words (iambic vs trochaic).

Although Canadian French 8-month-olds could extract disyllables from Canadian and European French utterances (Polka & Sundara, 2012), European French infants were not found to achieve this, even in their native dialect, until 16 months (Nazzi, Iakimova, Bertoncini, Frédonie, & Alcantara, 2006). Speech style may be a critical difference between studies: the European French stimuli (Nazzi et al.) "were less infant-directed […] produced with a faster speech rate, lower pitch, and smaller pitch excursions" (Nazzi, Mersad, Sundara, Iakimova, & Polka, 2014). The importance of IDS style for segmentation was highlighted by Floccia et al. (2016): reporting 13 studies of British English infants, from 8 to 10.5 months, they failed to replicate American English findings in all but one study. The positive segmentation effect was found with 10.5-month-old infants listening to speech in an "exaggerated IDS" style, which Floccia et al. suggested was closer to natural American English IDS.

Segmentation differences between (both French- and English-learning) North American and European infants thus may relate to contrasts in the typical style of IDS to which they are regularly exposed, rather than merely arising from between-study methodological differences. Consequently, Floccia et al. suggested that exaggerated North American IDS may contribute to higher vocabulary scores for American compared to British children between one and two years (Hamilton, Plunkett, & Schafer, 2000). This would accord with behavioural and electrophysiological studies showing a link between performance on segmentation tasks in the first year and language proficiency later in childhood (Junge & Cutler, 2014; Junge, Cutler, & Hagoort, 2012; Kooijman, Junge, Johnson, Hagoort, & Cutler, 2013; Newman, Ratner, Jusczyk, Jusczyk, & Dow, 2006). Furthermore, the relationship between pre-12-months segmentation outcomes and language proficiency is specific: early performance on language discrimination tasks is not predictive of vocabulary in 2-year-olds, and early segmentation performance predicts language scores, but not general cognitive abilities, between 4 and 6 years of age (Newman, Ratner, Jusczyk, Jusczyk, & Dow, 2006).

In addition to metrical segmentation, there have been demonstrations of infant sensitivity to a range of non-lexical cues exploited by adults in some listening contexts. Infants as young as 10 months can use prosodic edge cues, in particular lengthening before and after phrase boundaries, to determine the location of word boundaries (Gout, Christophe, & Morgan, 2004): both 10- and 13-month old children showed a familiarity response to words such as *paper* over phonologically-parallel

sequences separated by a phrase boundary, e.g., *pay performs*. Decoarticulation, other allophonic cues and phonotactics are all utilized by 8-month-olds (Johnson & Jusczyk, 2001; Mattys & Jusczyk, 2001a, 2001b), who also have a preference for words with onset consonants rather than vowels (Mattys & Jusczyk, 2001a).

An influential line of research considers infants' ability to extract statistical regularities from speech and use them to infer the location of word boundaries. Saffran et al. (1996) familiarized 8-month-olds to artificial languages composed of four nonwords (*pabiku*, *golatu*, *tibudo*, *daropi*), finding that they subsequently listened longer to the part-words *tudaro* and *pigola* rather than the words themselves. Although the part-words occurred in the language stream, across the boundaries of the words themselves (e.g., *daropigolatu*), the infants showed sensitivity to the frequencies of co-occurrence of successive syllables, preferring the relatively infrequent sequences. Subsequent studies demonstrated not merely sensitivity to overall sequential frequencies, but to their transitional probabilities, i.e., (frequency of syllable B following syllable A)/(frequency of syllable A).

The use of such statistical regularities is often contrasted with exploitation of other non-lexical cues, such as lexical stress (Johnson & Jusczyk, 2001). Given that non-lexical cues are essentially stochastic rather than deterministic in nature, however, the development of segmentation heuristics, whether framed as *statistical* or not, presumably relies on extraction of regularities from the signal. Thus, rather than using categorically distinct segmentation strategies, infants can be seen as extracting regularities from speech at different levels of abstraction. To exploit native stress patterns, English infants must classify syllables into strong and weak, and generalize that most words begin with a strong syllable. Phonotactic segmentation requires a generalization at the phonemic level. Allophonic cues such as decoarticulation require recognition of sub-phonemic patterns (i.e., the same phoneme realized differently in boundary and word-internal contexts). Most specifically, infants can utilize knowledge of individual words.

Thus, infants' exploitation of segmentation cues typically progresses from the more general to the more specific, whilst – at any particular stage of development – making use of whatever knowledge of contingent regularities they have thus far extracted.

7.3 Non-native segmentation

Visiting another country and hearing an unfamiliar language, who has not had the impression that native speakers are talking unusually quickly? Cutler (2012) calls this – often illusory – impression the

"gabbling foreigner" effect, and it arises from our inability to process the unfamiliar stream of sounds into discrete words, in contrast with our native language.

First and second language acquisition appear similar with respect to implicit segmentation through word recognition, but somewhat distinct regarding explicit strategies. As for infants (Bortfeld et al., 2005), word knowledge appears to be exploited for L2 segmentation as soon as acquired. Thus, Hungarian native speakers, even at low levels of English L2 proficiency, used segmentation-by-lexical-subtraction when listening to English (White, Melhorn, & Mattys, 2010). Lexical segmentation effects, similar to native speakers, have been observed for Japanese and Spanish speakers of L2 English (Sanders, 2003); however, an electrophysiological signature of segmentation in native speakers, the word-onset-associated N100, was found to be absent even for competent Japanese speakers of L2 English.

Cutler and Otake (1994) reflected that fluent bilinguals may operate effectively in two languages whilst manifesting segmentation biases specific to one or other language. Thus, implicit segmentation may sometimes mask differences in explicit strategies. Weber and Cutler (2006) found that proficient German speakers of L2 English could use English phonotactics for segmentation, whilst showing some persistence of German-derived regularities. Similarly, competent Japanese and Spanish speakers of L2 English used stress for segmentation purposes where lexical information was not available, but their specific application of stress depended on native language characteristics (Sanders, Neville, & Woldorff, 2002). Similarly, Tremblay and Spinelli (2014) found that competent English speakers of L2 French exploited distributional information like native French speakers for distinguishing word-initial and liaison consonants, but also showed sensitivity to L1-relevant durational cues. Thus, the use of non-lexical segmentation cues – stress, phonotactics, timing, allophony – all show influences of both L1 and L2 experience.

8 Summary and outlook

First and second language learners' segmentation behaviour demonstrates the power of lexical knowledge, as described earlier for adult listeners. Where segmentation can be achieved implicitly through word recognition, there may be minimal use of non-lexical cues. However, explicit segmentation strategies, drawing on a range of generalizations about native language sound patterns, are critical in many speech contexts. Adult listeners faced with sub-optimal listening conditions, from noise and degraded input to ambiguity and imperfect lexical knowledge, must call upon segmental and prosodic segmentation cues relevant to their native or non-native languages.

Likewise, infants will exploit whatever regularities they can infer from their linguistic experience, starting with the most general, to facilitate identification of word edges and expand their vocabulary, thus gradually moving to an implicit mode of segmentation.

Neuroscientific methods offer the prospect of new perspectives on some long-running debates in psycholinguistics in general, and speech segmentation in particular. A fully developed theory of segmentation should specify the units of representation – phonemes, syllables, words, etc. – that are required to extract structure from the signal, and should account for the integration and relative influence of signal-derived and knowledge-derived cues to that structure. Cognitive neuroscience has shown potential for insights, complementary to the well-established results of behavioural psycholinguistics, on both of these key questions. Firstly, electroencephalography (EEG) studies have measured event-related potentials (ERPs) in listeners which correspond to structure in auditory linguistic input, notably – with regard to word segmentation – the N100 (Sanders, Newport, & Neville, 2002). Secondly, many recent studies have demonstrated the entrainment of endogenous neural oscillations to the speech signal (see Peelle & Davis, 2012, for a review), an acoustically-driven effect that is neither special to speech nor to humans (Steinschneider, Nourski, & Fishman, 2013), but nonetheless has generated much theoretical interest regarding speech perception. In particular, phase-resetting of theta oscillations (in the 4-8Hz range) tracks the structure of the speech amplitude envelope, which has been taken as a rough proxy for syllable-level organisation (but see Cummins, 2012, regarding the limitations of this parallel). Furthermore, some studies have shown an enhancement of theta-range phase-locking for intelligible compared to unintelligible speech (e.g., Peelle, Gross, & Davis, 2013), whilst Ding, Chatterjee and Simon (2014) found entrainment in the delta range to be predictive of listeners' comprehension of the signal. Such interactions between acoustic and linguistic processing suggest the potential for neuroscience methods to further illuminate how knowledge-based prediction is integrated with the range of acoustic and sub-lexical segmentation cues described above.

More generally, segmentation research has yet to establish which cues apply regardless of specific linguistic background, information that may enhance understanding of language evolution and change. The role of broader cognitive processes in language processing will be informed by studies of individual differences in segmentation tasks. Finally, edging out of the laboratory to explore the dynamic negotiation of support for segmentation in natural conversation is a challenge yet to be fully confronted.

9 References

Albin, D. D., & Echols, C. H. (1996). Stressed and word-final syllables in infant-directed speech. *Infant Behavior and Development*, *19*(4), 401–418.

Anderson, J., & Jones, C. (1974). Three theses concerning phonological representations. *Journal of Linguistics*, *10*(01), 1. https://doi.org/10.1017/S0022226700003972

Berkovits, R. (1991). The effect of speaking rate on evidence for utterance-final lengthening. *Phonetica*, *48*(1), 57–66. https://doi.org/10.1159/000261871

Bolinger, D. (1964). Intonation as a universal. In *Proceedings of the 5th Congress of Phonetics, Cambridge 1962* (pp. 833–848).

Bortfeld, H., Morgan, J. L., Golinkoff, R. M., & Rathbun, K. (2005). Mommy and me familiar names help launch babies into speech-stream segmentation. *Psychological Science*, *16*(4), 298–304.

Bowers, J. S., Davis, C. J., Mattys, S. L., Damian, M. F., & Hanley, D. (2009). The activation of embedded words in spoken word identification is robust but constrained: Evidence from the picture-word interference paradigm. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(5), 1585.

Brent, M. R. (1997). Toward a unified model of lexical acquisition and lexical access. *Journal of Psycholinguistic Research*, *26*(3), 363–375.

Brent, M. R., & Siskind, J. M. (2001). The role of exposure to isolated words in early vocabulary development. *Cognition*, *81*(2), B33–B44.

Cole, R. A., Jakimik, J., & Cooper, W. E. (1980). Segmenting speech into words. *The Journal of the Acoustical Society of America*, *67*(4), 1323–1332.

Content, A., Kearns, R. K., & Frauenfelder, U. H. (2001). Boundaries versus onsets in syllabic segmentation. *Journal of Memory and Language*, *45*(2), 177–199.

Content, A., Meunier, C., Kearns, R. K., & Frauenfelder, U. H. (2001). Sequence detection in pseudowords in French: Where is the syllable effect? *Language and Cognitive Processes*, *16*(5–6), 609–636.

Cristia, A. (2013). Input to language: The phonetics and perception of infant-directed speech. *Language and Linguistics Compass*, *7*(3), 157–170.

Cummins, F. (2012). Oscillators and syllables: a cautionary note. *Frontiers in Psychology*, *3*. https://doi.org/10.3389/fpsyg.2012.00364

Cunillera, T., Càmara, E., Laine, M., & Rodríguez-Fornells, A. (2010). Words as anchors: known words facilitate statistical learning. *Experimental Psychology*, *57*(2), 134–141. https://doi.org/10.1027/1618-3169/a000017

Cunillera, T., Laine, M., & Rodríguez-Fornells, A. (2016). Headstart for speech segmentation: a neural signature for the anchor word effect. *Neuropsychologia*, *82*, 189–199. https://doi.org/10.1016/j.neuropsychologia.2016.01.011

Cutler, A. (2012). *Native Listening: Language Experience and the Recognition of Spoken Words*. MIT Press.

Cutler, A., & Butterfield, S. (1992). Rhythmic cues to speech segmentation: Evidence from juncture misperception. *Journal of Memory and Language*, *31*(2), 218–236.

Cutler, A., & Carter, D. M. (1987). The predominance of strong initial syllables in the English vocabulary. *Computer Speech & Language*, *2*(3), 133–142.

Cutler, A., & Foss, D. J. (1977). On the role of sentence stress in sentence processing. *Language and Speech*, *20*(1), 1–10.

Cutler, A., McQueen, J. M., Norris, D., & Somejuan, A. (2001). The roll of the silly ball. In E. Dupoux (Ed.), *Language, brain and cognitive development: Essays in honor of Jacques Mehler* (pp. 181–194). Cambridge, MA: MIT Press.

Cutler, A., Mehler, J., Norris, D., & Segui, J. (1986). The syllable's differing role in the segmentation of French and English. *Journal of Memory and Language*, *25*(4), 385–400.

Cutler, A., & Norris, D. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance*, *14*(1), 113.

Cutler, A., & Otake, T. (1994). Mora or phoneme? Further evidence for language-specific listening. *Journal of Memory and Language*, *33*(6), 824.

Dahan, D., & Brent, M. R. (1999). On the discovery of novel wordlike units from utterances: an artificial-language study with implications for native-language acquisition. *Journal of Experimental Psychology: General*, *128*(2), 165–185.

Davis, M. H., Marslen-Wilson, W. D., & Gaskell, M. G. (2002). Leading up the lexical garden path: Segmentation and ambiguity in spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *28*(1), 218–244. https://doi.org/10.1037//0096-1523.28.1.218

Dilley, L., Shattuck-Hufnagel, S., & Ostendorf, M. (1996). Glottalization of word-initial vowels as a function of prosodic structure. *Journal of Phonetics*, *24*(4), 423–444.

Ding, N., Chatterjee, M., & Simon, J. Z. (2014). Robust cortical entrainment to the speech envelope relies on the spectro-temporal fine structure. *Neuroimage*, *88*, 41–46.

Dumay, N., Frauenfelder, U. H., & Content, A. (2002). The role of the syllable in lexical segmentation in french: word-spotting data. *Brain and Language*, *81*(1–3), 144–161. https://doi.org/10.1006/brln.2001.2513

Dunst, C., Gorman, E., & Hamby, D. (2012). Preference for infant-directed speech in preverbal young children. *Center for Early Literacy Learning*, *5*(1), 1–13.

Fear, B. D., Cutler, A., & Butterfield, S. (1995). The strong/weak syllable distinction in English. *The Journal of the Acoustical Society of America*, *97*(3), 1893–1904.

Fernald, A., & McRoberts, G. (1996). Prosodic bootstrapping: A critical analysis of the argument and the evidence. *Signal to Syntax: Bootstrapping from Speech to Grammar in Early Acquisition*, 365–388.

Fernald, A., Taeschner, T., Dunn, J., Papousek, M., de Boysson-Bardies, B., & Fukui, I. (1989). A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants. *Journal of Child Language*, *16*(03), 477–501.

Fernandes, T., Ventura, P., & Kolinsky, R. (2007). Statistical information and coarticulation as cues to word boundaries: A matter of signal quality. *Perception & Psychophysics*, *69*(6), 856–864.

Finn, A. S., & Hudson Kam, C. L. (2008). The curse of knowledge: First language knowledge impairs adult learners' use of novel statistics for word segmentation. *Cognition*, *108*(2), 477–499. https://doi.org/10.1016/j.cognition.2008.04.002

Fisher, C., & Tokura, H. (1996). Acoustic cues to grammatical structure in infant-directed speech: cross-linguistic evidence. *Child Development*, *67*(6), 3192–3218. https://doi.org/10.1111/j.1467-8624.1996.tb01909.x

Floccia, C., Keren-Portnoy, T., DePaolis, R., Duffy, H., Delle Luche, C., Durrant, S., … Vihman, M. (2016). British English infants segment words only with exaggerated infant-directed speech stimuli. *Cognition*, *148*, 1–9. https://doi.org/10.1016/j.cognition.2015.12.004

Fougeron, C., & Keating, P. A. (1997). Articulatory strengthening at edges of prosodic domains. *The Journal of the Acoustical Society of America*, *101*(6), 3728–3740.

Gout, A., Christophe, A., & Morgan, J. L. (2004). Phonological phrase boundaries constrain lexical access II. Infant data. *Journal of Memory and Language*, *51*(4), 548–567. https://doi.org/10.1016/j.jml.2004.07.002

Gow, D. W., & Gordon, P. C. (1995). Lexical and prelexical influences on word segmentation: Evidence from priming. *Journal of Experimental Psychology: Human Perception and Performance*, *21*(2), 344–359.

Hamilton, A., Plunkett, K., & Schafer, G. (2000). Infant vocabulary development assessed with a British communicative development inventory. *Journal of Child Language*, *27*(03), 689–705. https://doi.org/null

Hanulíková, A., McQueen, J. M., & Mitterer, H. (2010). Possible words and fixed stress in the segmentation of Slovak speech. *The Quarterly Journal of Experimental Psychology*, *63*(3), 555–579. https://doi.org/10.1080/17470210903038958

Harrington, J., Watson, G., & Cooper, M. (1989). Word boundary detection in broad class and phoneme strings. *Computer Speech & Language*, *3*(4), 367–382.

Huttenlocher, D. P., & Zue, V. W. (1983). Phonotactic and lexical constraints in speech recognition. In *Proceedings of the Third AAAI Conference on Artificial Intelligence* (pp. 172–176). AAAI Press.

Johnson, E. K., & Jusczyk, P. W. (2001). Word segmentation by 8-month-olds: When speech cues count more than statistics. *Journal of Memory and Language*, *44*(4), 548–567.

Junge, C., & Cutler, A. (2014). Early word recognition and later language skills. *Brain Sciences*, *4*(4), 532–559.

Junge, C., Cutler, A., & Hagoort, P. (2012). Electrophysiological evidence of early word learning. *Neuropsychologia*, *50*(14), 3702–3712. https://doi.org/10.1016/j.neuropsychologia.2012.10.012

Jusczyk, P. W., Houston, D. M., & Newsome, M. (1999). The beginnings of word segmentation in English-learning infants. *Cognitive Psychology*, *39*(3), 159–207.

Keating, P., Cho, T., Fougeron, C., & Hsu, C.-S. (2004). Domain-initial articulatory strengthening in four languages. In J. Local, R. Ogden, & R. Temple (Eds.), *Phonetic Interpretation: Papers in Laboratory Phonology VI* (pp. 143–161). Cambridge: Cambridge University Press.

Kim, D., Stephens, J. D., & Pitt, M. A. (2012). How does context play a part in splitting words apart? Production and perception of word boundaries in casual speech. *Journal of Memory and Language*, *66*(4), 509–529.

Kooijman, V., Junge, C., Johnson, E. K., Hagoort, P., & Cutler, A. (2013). Predictive brain signals of linguistic development. *Frontiers in Psychology*, *4*, 25. https://doi.org/10.3389/fpsyg.2013.00025

Ladd, D. R. (2008). *Intonational phonology*. Cambridge: Cambridge University Press.

Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&H theory. In *Speech Production and Speech Modelling* (pp. 403–439). Springer.

Mattys, S. L. (2004). Stress versus coarticulation: toward an integrated approach to explicit speech segmentation. *Journal of Experimental Psychology: Human Perception and Performance*, *30*(2), 397.

Mattys, S. L., & Bortfeld, H. (2015). Speech segmentation. In G. Gaskell & J. Mirkovic (Eds.), *Speech Perception and Spoken Word Recognition*. Taylor and Francis.

Mattys, S. L., & Jusczyk, P. W. (2001a). Do infants segment words or recurring contiguous patterns? *Journal of Experimental Psychology: Human Perception and Performance*, *27*(3), 644–655. https://doi.org/10.1037/0096-1523.27.3.644

Mattys, S. L., & Jusczyk, P. W. (2001b). Phonotactic cues for segmentation of fluent speech by infants. *Cognition*, *78*(2), 91–121.

Mattys, S. L., & Melhorn, J. F. (2005). How do syllables contribute to the perception of spoken English? Insight from the migration paradigm. *Language and Speech*, *48*(2), 223–252.

Mattys, S. L., Melhorn, J. F., & White, L. (2007). Effects of syntactic expectations on speech segmentation. *Journal of Experimental Psychology: Human Perception and Performance*, *33*(4), 960–977. https://doi.org/10.1037/0096-1523.33.4.960

Mattys, S. L., White, L., & Melhorn, J. F. (2005). Integration of multiple speech segmentation cues: a hierarchical framework. *Journal of Experimental Psychology: General*, *134*(4), 477–500. https://doi.org/10.1037/0096-3445.134.4.477

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18*(1), 1–86.

McQueen, J. M. (1998). Segmentation of continuous speech using phonotactics. *Journal of Memory and Language*, *39*(1), 21–46.

McQueen, J. M. (2005). Speech perception. In *The Handbook of Cognition* (pp. 255–275). London: Sage Publications.

McQueen, J. M., Cutler, A., Briscoe, T., & Norris, D. (1995). Models of continuous speech recognition and the contents of the vocabulary. *Language and Cognitive Processes*, *10*(3–4), 309–331.

McQueen, J. M., Norris, D., & Cutler, A. (1994). Competition in spoken word recognition: Spotting words in other words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(3), 621.

Mehler, J., Dommergues, J. Y., Frauenfelder, U., & Segui, J. (1981). The syllable's role in speech segmentation. *Journal of Verbal Learning and Verbal Behavior*, *20*(3), 298–305.

Morgan, J. L., & Demuth, K. (1996). *Signal to syntax: Bootstrapping from speech to grammar in early acquisition*. Psychology Press.

Mueller, J. L., Bahlmann, J., & Friederici, A. D. (2008). The role of pause cues in language learning: The emergence of event-related potentials related to sequence processing. *Journal of Cognitive Neuroscience*, *20*(5), 892–905.

Nakatani, L. H., & O'Connor-Dukes, K. (1980). Phonetic parsing cues for word perception. *Unpublished Manuscript. Murray Hill, NJ: Bell Laboratories*.

Nazzi, T., Iakimova, G., Bertoncini, J., Frédonie, S., & Alcantara, C. (2006). Early segmentation of fluent speech by infants acquiring French: Emerging evidence for crosslinguistic differences. *Journal of Memory and Language*, *54*(3), 283–299.

Nazzi, T., Mersad, K., Sundara, M., Iakimova, G., & Polka, L. (2014). Early word segmentation in infants acquiring Parisian French: task-dependent and dialect-specific aspects. *Journal of Child Language*, *41*(03), 600–633. https://doi.org/10.1017/S0305000913000111

Nelson, D. G. K., Hirsh-Pasek, K., Jusczyk, P. W., & Cassidy, K. W. (1989). How the prosodic cues in motherese might assist language learning. *Journal of Child Language*, *16*(01), 55–68.

Nespor, M., Shukla, M., & Mehler, J. (2011). Stress-timed vs. syllable-timed languages. *The Blackwell Companion to Phonology*, *2*, 1147–1159.

Newman, R. S., Ratner, N. B., Jusczyk, A. M., Jusczyk, P. W., & Dow, K. A. (2006). Infants' early ability to segment the conversational speech signal predicts later language development: a retrospective analysis. *Developmental Psychology*, *42*(4), 643.

Newman, R. S., Sawusch, J. R., & Wunnenberg, T. (2011). Cues and cue interactions in segmenting words in fluent speech. *Journal of Memory and Language*, *64*(4), 460–476. https://doi.org/10.1016/j.jml.2010.11.004

Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition*, *52*(3), 189–234.

Norris, D., & Cutler, A. (1985). Juncture detection. *Linguistics*, *23*(5), 689–706.

Norris, D., & McQueen, J. M. (2008). Shortlist B: a Bayesian model of continuous speech recognition. *Psychological Review*, *115*(2), 357.

Norris, D., McQueen, J. M., & Cutler, A. (1995). Competition and segmentation in spoken-word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(5), 1209.

Norris, D., McQueen, J. M., Cutler, A., & Butterfield, S. (1997). The possible-word constraint in the segmentation of continuous speech. *Cognitive Psychology*, *34*(3), 191–243.

Öhman, S. E. (1966). Coarticulation in VCV utterances: Spectrographic measurements. *The Journal of the Acoustical Society of America*, *39*(1), 151–168.

Oller, D. K. (1973). The effect of position in utterance on speech segment duration in English. *The Journal of the Acoustical Society of America*, *54*(5), 1235–1247.

Ordin, M., Polyanskaya, L., Laka, I., & Nespor, M. (2017). Cross-linguistic differences in the use of durational cues for the segmentation of a novel language. *Memory & Cognition*, 1–14. https://doi.org/10.3758/s13421-017-0700-9

Peelle, J. E., & Davis, M. H. (2012). Neural oscillations carry speech rhythm through to comprehension. *Frontiers in Psychology*, *3*. https://doi.org/10.3389/fpsyg.2012.00320

Peelle, J. E., Gross, J., & Davis, M. H. (2013). Phase-locked responses to speech in human auditory cortex are enhanced during comprehension. *Cerebral Cortex*, *23*(6), 1378–1387. https://doi.org/10.1093/cercor/bhs118

Pelucchi, B., Hay, J. F., & Saffran, J. R. (2009). Statistical learning in a natural language by 8-month-old infants. *Child Development*, *80*(3), 674–685. https://doi.org/10.1111/j.1467-8624.2009.01290.x

Peña, M., Bonatti, L. L., Nespor, M., & Mehler, J. (2002). Signal-driven computations in speech processing. *Science*, *298*(5593), 604–607.

Polka, L., & Sundara, M. (2012). Word segmentation in monolingual infants acquiring Canadian English and Canadian French: Native language, cross-dialect, and cross-language comparisons. *Infancy*, *17*(2), 198–232. https://doi.org/10.1111/j.1532-7078.2011.00075.x

Pompino-Marschall, B., & Żygis, M. (2010). Glottal marking of vowel-initial words in German. *ZAS Papers in Linguistics*, *52*, 1–17.

Price, P. J., Ostendorf, M., Shattuck-Hufnagel, S., & Fong, C. (1991). The use of prosody in syntactic disambiguation. *The Journal of the Acoustical Society of America*, *90*(6), 2956–2970.

Reinisch, E., Jesse, A., & McQueen, J. M. (2011). Speaking Rate Affects the Perception of Duration as a Suprasegmental Lexical-stress Cue. *Language and Speech*, *54*(2), 147–165. https://doi.org/10.1177/0023830910397489

Rossi, S., Jürgenson, I. B., Hanulíková, A., Telkemeyer, S., Wartenburger, I., & Obrig, H. (2011). Implicit processing of phonotactic cues: evidence from electrophysiological and vascular responses. *Journal of Cognitive Neuroscience*, *23*(7), 1752–1764.

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language*, *35*(4), 606–621.

Salverda, A. P., Dahan, D., & McQueen, J. M. (2003). The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension. *Cognition*, *90*(1), 51–89. https://doi.org/10.1016/S0010-0277(03)00139-2

Sanders, L. D. (2003). An ERP study of continuous speech processing: II. Segmentation, semantics, and syntax in non-native speakers. *Cognitive Brain Research*, *15*(3), 214–227.

Sanders, L. D., Neville, H. J., & Woldorff, M. G. (2002). Speech segmentation by native and non-native speakers: The use of lexical, syntactic, and stress-pattern cues. *Journal of Speech, Language, and Hearing Research*, *45*(3), 519–530.

Sanders, L. D., Newport, E. L., & Neville, H. J. (2002). Segmenting nonsense: an event-related potential index of perceived onsets in continuous speech. *Nature Neuroscience*, *5*(7), 700–703.

Sebastián-Gallés, N., Dupoux, E., Segui, J., & Mehler, J. (1992). Contrasting syllabic effects in Catalan and Spanish. *Journal of Memory and Language*, *31*(1), 18–32.

Shatzman, K. B., & McQueen, J. M. (2006). Segment duration as a cue to word boundaries in spoken-word recognition. *Perception & Psychophysics*, *68*(1), 1–16.

Shillcock, R. C. (1990). Lexical hypotheses in continuous speech. In G. T. M. Altmann (Ed.), *Cognitive models of speech processing: Psycholinguistic and computational perspectives* (pp. 24–49). Cambridge, MA: MIT Press.

Shukla, M., Nespor, M., & Mehler, J. (2007). An interaction between prosody and statistics in the segmentation of fluent speech. *Cognitive Psychology*, *54*(1), 1–32. https://doi.org/10.1016/j.cogpsych.2006.04.002

Skoruppa, K., Nevins, A., Gillard, A., & Rosen, S. (2015). The role of vowel phonotactics in native speech segmentation. *Journal of Phonetics*, *49*, 67–76.

Smiljanić, R., & Bradlow, A. R. (2008). Temporal organization of English clear and conversational speech. *The Journal of the Acoustical Society of America*, *124*(5), 3171. https://doi.org/10.1121/1.2990712

Smiljanić, R., & Bradlow, A. R. (2009). Speaking and hearing clearly: talker and listener factors in speaking style changes. *Language and Linguistics Compass*, *3*(1), 236–264. https://doi.org/10.1111/j.1749-818X.2008.00112.x

Spinelli, E., Grimault, N., Meunier, F., & Welby, P. (2010). An intonational cue to word segmentation in phonemically identical sequences. *Attention, Perception, & Psychophysics*, *72*(3), 775–787.

Steinschneider, M., Nourski, K. V., & Fishman, Y. I. (2013). Representation of speech in human auditory cortex: Is it special? *Hearing Research*, *305*, 57–73. https://doi.org/10.1016/j.heares.2013.05.013

Suomi, K., McQueen, J. M., & Cutler, A. (1997). Vowel harmony and speech segmentation in Finnish. *Journal of Memory and Language*, *36*(3), 422–444.

Tabossi, P., Collina, S., Mazzetti, M., & Zoppello, M. (2000). Syllables in the processing of spoken Italian. *Journal of Experimental Psychology: Human Perception and Performance*, *26*(2), 758.

Thiessen, E. D., Hill, E. A., & Saffran, J. R. (2005). Infant-Directed Speech Facilitates Word Segmentation. *Infancy*, *7*(1), 53–71. https://doi.org/10.1207/s15327078in0701_5

Tremblay, A., & Spinelli, E. (2014). English listeners' use of distributional and acoustic-phonetic cues to liaison in French: Evidence from eye movements. *Language and Speech*, *57*(3), 310–337.

Turk, A. E., & White, L. (1999). Structural influences on accentual lengthening in English. *Journal of Phonetics*, *27*(2), 171–206. https://doi.org/10.1006/jpho.1999.0093

Tyler, M. D., & Cutler, A. (2009). Cross-language differences in cue use for speech segmentation. *The Journal of the Acoustical Society of America*, *126*(1), 367–376.

Vroomen, J., & de Gelder, B. (1995). Metrical segmentation and lexical inhibition in spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *21*(1), 98–108. https://doi.org/10.1037/0096-1523.21.1.98

Vroomen, J., & De Gelder, B. (1997). Activation of embedded words in spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *23*(3), 710.

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*(2), 133–149.

Weber, A., & Cutler, A. (2006). First-language phonotactics in second-language listening. *The Journal of the Acoustical Society of America*, *119*(1), 597–607.

White, L. (2014). Communicative function and prosodic form in speech timing. *Speech Communication*, *63–64*, 38–54. https://doi.org/10.1016/j.specom.2014.04.003

White, L., Mattys, S. L., Stefansdottir, L., & Jones, V. (2015). Beating the bounds: Localized timing cues to word segmentation. *The Journal of the Acoustical Society of America*, *138*(2), 1214–1220. https://doi.org/10.1121/1.4927409

White, L., Mattys, S. L., & Wiget, L. (2012a). Language categorization by adults is based on sensitivity to durational cues, not rhythm class. *Journal of Memory and Language*, *66*(4), 665–679. https://doi.org/10.1016/j.jml.2011.12.010

White, L., Mattys, S. L., & Wiget, L. (2012b). Segmentation cues in conversational speech: robust semantics and fragile phonotactics. *Frontiers in Psychology*, *3:375*. https://doi.org/10.3389/fpsyg.2012.00375

White, L., Melhorn, J. F., & Mattys, S. L. (2010). Segmentation by lexical subtraction in Hungarian speakers of second-language English. *The Quarterly Journal of Experimental Psychology*, *63*(3), 544–554. https://doi.org/10.1080/17470210903006971

White, L., & Turk, A. E. (2010). English words on the Procrustean bed: Polysyllabic shortening reconsidered. *Journal of Phonetics*, *38*(3), 459–471. https://doi.org/10.1006/jpho.1999.0093

Wightman, C. W., Shattuck-Hufnagel, S., Ostendorf, M., & Price, P. J. (1992). Segmental durations in the vicinity of prosodic phrase boundaries. *The Journal of the Acoustical Society of America*, *91*(3), 1707–1717.