4

SPEECH SEGMENTATION

Sven L. Mattys and Heather Bortfeld

Introduction

Inspection of a speech waveform does not reveal clear correlates of what the listener perceives as word boundaries. Yet the absence of word boundary markers hardly poses a problem for listeners, as the subjective experience of speech is not of continuity but, rather, of discreteness – a sequence of individual words. This chapter is concerned with the perceptual and cognitive mechanisms underlying conversion of continuity into discreteness, namely speech segmentation.

Or so goes the story. Over the past 20 years, the concept of segmentation has been challenged on two fronts. First, how truly continuous is speech? Could the presence of clear boundary markers (e.g., pauses) be underestimated? This question is of particular importance for early stages of word learning, as infant-directed speech is known to have communication-oriented characteristics, some of which could guard against, or at least mitigate, continuity. Second, even if speech requires segmentation, to what extent does it need to be investigated separately from word recognition? Do we need to construe segmentation as a battery of cues and mechanisms, or is it an emergent property of lexical activation, an epiphenomenon of lexical competition?

These questions motivate the structure of this chapter. We argue that much depends on whether the language system has achieved a mature, steady state, is in a process of development, or is facing suboptimal listening conditions.

Segmentation in the mature system

For adult native speakers, segmentation consists of correctly mapping sound strings onto word representations, a process undermined by the temporal distribution of sounds and by lexical embeddedness (Cutler, 2012). The fact that the distribution

of auditory information is time-bound and transient means that lexical activation must start before an entire word has been heard, resulting in multiple simultaneous activations. Embeddedness only exacerbates the problem. Some 84% of polysyllabic English words have at least one shorter word embedded in them ("captain" contains "cap"); indeed, most short words are embedded in longer ones (McQueen et al., 1995). Spurious words are also found across word boundaries ("belt" in "bell tower"). Corpus data in English reveal that more than a third of all word boundaries contain at least one embedded word (Roach et al., 1993).

Mechanisms underlying speech segmentation have been investigated from two standpoints: as an emergent property of lexical competition and contextual expectations (lexical and post-lexical phenomena) and as the product of a constellation of segmental and prosodic cues (non-lexical phenomena).

Segmentation as a lexical (and post-lexical) phenomenon

Word and world knowledge often constrain segmentation. For instance, "The goose trader needs a licence" might contain the spurious words "goo", "stray", "tray", "knee", "lie", "ice", in addition to the intended words, but only the intended words make the utterance lexically and semantically viable. Activation of the spurious words is therefore likely to be weak, especially in later parts of the sentence. Likewise, although cases of fully overlapping segmentation alternatives exist (e.g., "known ocean" vs, "no notion"), they are rare and usually get disambiguated by syntactic and/or semantic context. Segmentation is also constrained by whether or not the outcome *could be* a real word. In competition-based models, the possible-word constraint (Norris et al., 1997) reduces activation of word candidates that create phonologically impossible adjacent words. For instance, in "black bird", "lack" will be disfavoured not only because the preceding "b" portion is not an existing English word but also because it violates English morphological rules.

Modeling work demonstrates good segmentation performance through multiple lexical activation and competition (e.g., TRACE: McClelland & Elman, 1986; Shortlist: Norris, 1994). On this view, all words in running speech are temporarily activated, with relative activation levels fluctuating as a function of fit with the unfolding input. The activation of spurious words, possibly high at times (e.g., "goo" in "goose trader"), eventually is terminated through direct inhibition by better fitting competitors (TRACE) or dies down due to lack of sensory confirmation (Shortlist). A lexically guided process is effective for new words too. For instance, Dahan and Brent (1999) showed that listeners familiarized with five-syllable sequences (e.g., "dobuneripo") and three-syllable sequences embedded in the longer ones (e.g., "neripo"), subsequently better remembered a sequence corresponding with the remainder sequence (e.g., "dobu") than one that did not (e.g., "dobune"). Such segmentation-by-lexical-subtraction process likewise boosts first- and second-language learning (Bortfeld et al., 2005; White et al., 2010).

Semantic and syntactic expectations also provide constraints. Kim and colleagues (2012) showed that the segmentation of a stimulus ("along") was significantly

determined by its sentential context ("along" in "try to get __" and "a long" in "it takes __ time"). Likewise, Mattys and associates (2007) found that segmentation of an ambiguous stimulus like "takespins" was influenced by whether the syntactic context suggested that the pivotal /s/ was a third-person inflection or the first phoneme of the object noun phrase.

Importantly, this knowledge-driven approach does not assign a specific computational status to segmentation, other than as the consequence of mechanisms associated with lexical competition and interactions with higher-order knowledge.

Segmentation as a non-lexical, cue-based, and probabilistic phenomenon

Although an effective mechanism in a majority of cases, lexical-driven segmentation fails in cases of multiple parses ("stray tissues"/"straight issues") and whenever new or unfamiliar words are encountered. Non-lexical cues fall within various categories. Some arise from distributional regularities within the lexicon (probabilistic phonotactics, lexical stress), and others operate independently (allophonic variations). Some are language-specific (generally those arising from regularities in the lexicon), and others are fairly constant across languages (some but not all rhythmic cues). Some align with actual word boundaries, and others contribute to word boundary detection only insofar as they align with phrase- or sentence-level boundaries (final lengthening). Because these categories are somewhat artificial and not mutually exclusive, we opt for a description based on whether non-lexical cues involve segmental, subsegmental, or suprasegmental regularities.

Segmental cues

Segmental cues refer to the probability that certain phonemes or sequences of phonemes align with word boundaries. For instance, the /j/ sound in English is more likely to be word-medial or word-final than word-initial. Similarly, sequences rarely found at the beginning of words (/mr/) are likely to be interpreted as straddling a word boundary (McQueen, 1998; Tagliapietra et al., 2009). These regularities are referred to as phonotactic cues. Since such regularities originate in the phonological structure of a language, sequences of phonemes perceived as legitimate word boundaries vary from language to language.

Rules about word-internal segmental harmony can also point to boundaries. In Finnish, for example, rules of harmony prevent certain categories of vowels from co-occurring within a word. Thus, assuming that vowel V_i cannot be found within the same word as vowel V_{ii}, a sequence like CV_iCV_{ii} could not constitute a single word; instead, it would have to include a word boundary somewhere between the two vowels (Suomi et al., 1997; Vroomen et al., 1998). Although they are weaker heuristics, segmental sandhis – a broad range of phonological changes that occur at morpheme or word boundaries – also cue segmentation. Liaison, for instance, is the addition of a segment between two words to avoid a non-permissible sequence

of sounds. For example, while "petit frère" (younger brother) is produced as /pətifʁɛʁ/, "petit ami" (boyfriend) is produced as /pətit̲ami/, where /t/ is inserted between the two words to avoid a clash between /i/ and /a/. Since the most common liaising phonemes in French are /t/, /z/, and /n/, the occurrence of those sounds in continuous speech could be used as cues to word boundaries (Tremblay & Spinelli, 2014). Note that liaison-driven segmentation may partly be accounted for by acoustic differences between liaison phonemes and their non-liaison equivalents (Spinelli et al., 2003).

While lab-based findings show that segmentation is aided by listeners' sensitivity to phonotactic regularities, they represent an idealized view of how words are produced in everyday speech. Because conversational speech is plagued with phoneme deletions (Schuppler et al., 2011), insertions (Warner & Weber, 2001) and alterations, such as contextual variants (Janse et al., 2007) or phonological assimilation (Ohala, 1990), phonotactic statistics derived from lexical databases or transcribed corpora are only approximate. Indeed, we will argue that such expectations form a relatively minor influence on segmentation.

Subsegmental cues

The acoustic realization of a given phoneme (allophonic variant) can change from utterance to utterance. Some of this variation is reliably associated with word boundaries. We already mentioned that liaison phonemes in French are acoustically distinguishable from non-liaison phonemes, such that their acoustic signature could cue word junctures. Subsegmental cues allow listeners to distinguish phrase pairs ("dernier onion – dernier rognon" in French; "di amanti – diamanti" in Italian; "grey tanker – great anchor" in English) through a variety of phonetic processes such as word-onset glottalization (/ʔ/anchor), word-onset aspiration (/^{h}/tanker) and devoicing (in /r/ in "nitrate"). Position-specific durational contrasts (longer/pai/ in "high pay][per month . . .", where][is a phrase boundary, than in "paper") and changes in articulatory strength (phonemes are more weakly coarticulated across word boundaries than within words) provide additional cues.

The advantage of these subsegmental cues is that they supply information about word boundaries where lexical information cannot. Empirical evidence for listeners' sensitivity to these cues is abundant (e.g., Christophe et al., 2004; Davis et al., 2002; Shatzman & McQueen, 2006). However, in everyday listening conditions, lexically ambiguous phrases (e.g., "nitrate" vs. "night rate") are seldom heard outside of a disambiguating sentential context (e.g., "The soil contained a high concentration of nitrate."), which makes the actual contribution of subsegmental cues perhaps less substantial than laboratory speech research suggests. Furthermore, the reliability of subsegmental cues for segmentation is difficult to quantify. There are substantial individual variations in how systematically and fully speakers realize allophonic variations. Moreover, allophonic variations differ markedly in their acoustic salience, some showing clear qualitative contrasts (glottalization) and others showing small quantitative contrasts (lengthening). Finally, while some allophonic variations are

chiefly language-general (final lengthening), most are language-specific (aspiration, glottalization). Thus, languages differ considerably in the number, type, and effectiveness of the sub-lexical cues they provide. As with segmental cues, subsegmental cues act as supplementary heuristics rather than deterministic boundary markers.

Suprasegmental cues

Suprasegmental cues are variations in duration, fundamental frequency, and amplitude that coincide with word boundaries either proximally or distally. Lengthening is found within the word, phrase, and sentence domains (Wightman et al., 1992). A significant challenge for a speech recognizer, however, is that lengthening has been documented at both the initial and the final edges of domains. However, lengthening within a syllable may be interpretable as a word onset as opposed to a word offset by taking the distribution of the durational effect into account (Monaghan et al., 2013; White, 2014).

Fundamental frequency (*F*0), the acoustic source of perceived pitch, shows a high degree of correlation with major syntactic boundaries (Vaissière, 1983) and hence with the boundaries of words at the edge of syntactic constituents. Pitch movements that specifically align with word boundaries, independent of syntactic boundaries, are more elusive and are often constrained by other factors. For instance, Welby (2007) showed that the early rise in *F*0 typically found at the left edge of French content words cues the beginning of a new word, but this effect is modulated by other suprasegmental cues (phoneme duration) and lexical frequency. The locus of pitch movements relative to word boundaries is disputed, even within a single language. Using an artificial language learning paradigm (Saffran et al., 1996a), Tyler and Cutler (2009) showed that French listeners used right-edge rather than left-edge pitch movements as word boundary cues. English listeners showed the opposite pattern, and Dutch listeners used both left- and right-edge pitch movements. Although discrepancies may be partly due to differences in materials, tasks, and methods of pitch manipulation, it is clear that *F*0 alone provides limited support for the detection of word boundaries (see Toro et al., 2009, for additional evidence).

The combination of suprasegmental cues provides a stronger heuristic, such as is the case for lexical stress, the accentuation of syllables within words. Languages differ in the suprasegmental features (*F*0, duration, intensity) contributing to stress and the position of typically stressed syllables. Within a language, however, stress can be a highly reliable cue for word boundaries. In so-called fixed-stress languages, where stress always falls in a particular position (fixed word–final stress in Hungarian), inference based on stress should lead to high word boundary detection. In free-stress languages (English, Dutch, Italian), where stress placement varies, predominant stress positions can be used as a probabilistic cue. For instance, the predominance of stress-initial words is exploited in English (Cutler & Norris, 1988) and in Dutch (Vroomen et al., 1996). Languages for which stress is not a contrastive feature might provide rhythmic cues tied to their own phonology (syllables for French, Banel & Bacri, 1997; morae for Japanese, Otake et al., 1993). Cutler (2012)

provides an extensive review of how listeners use rhythm in general and lexical stress in particular to infer word boundaries.

While these suprasegmental cues can be characterized as proximal prosody, distal prosody leads to grouping and segmentation of speech materials several syllables after the prosodic cues themselves. Inspired by research on auditory grouping, Dilley and McAuley (2008) created auditory sequences beginning with two stress-initial words and ending with four syllables that could form words in more than one way. They then manipulated the duration and/or the *F*0 of both of the first five syllables in order to create a rhythmic pattern to induce either a weak-strong-weak (WSW) or strong-weak-strong (SWS) grouping in the final un-manipulated syllables. When participants were asked to report the last word of the sequence, they produced more monosyllabic words in the SWS-inducing condition and more disyllabic words in the WSW-inducing condition. The manipulation of either duration, or *F*0, worked, and the combination of both cues yielded an even larger effect. These results and others (Brown et al., 2011) are groundbreaking because they reveal anticipatory segmentation mechanisms that generate place holders for words not yet heard. It is precisely those types of mechanisms that could disambiguate sentence-final phrases like "night rate" and "nitrate", alongside the sentential context and the acoustic realisation of the test words themselves.

Segmentation as a multi-cue integration phenomenon

The investigation of cues in isolation highlights sources of information that the perceptual system can use. How they are used – if at all – is another matter. Likewise, computational models offer suggestions about how segmentation can be achieved parsimoniously, not necessarily how segmentation actually happens. For instance, Cairns and colleagues (1997) devised a neural network that correctly detected a third of the word boundaries in the London-Lund corpus of conversational English relying solely on phonotactic regularities. In contrast, TRACE assumes a pre-existing lexicon, entirely eschewing the involvement of non-lexical cues by making segmentation a by-product of lexical competition and selection (Frauenfelder & Peeters, 1990). Perhaps as a compromise, Shortlist (Norris, 1994) relies on lexically driven segmentation (per TRACE) but further fine-tunes it via statistical regularities (lexical stress, phonotactics) and corpus allophones (Shortlist B, Norris & McQueen, 2008). When between-word transitional probabilities are taken into account, the performance of computational models increases even further (Goldwater et al., 2009).

Although cues are likely to converge rather than diverge, pitting cues against one another has revealed much about their relative weights. Drawing on a battery of perceptual experiments, Mattys and colleagues (2005, 2007) demonstrated that listeners rely on lexical, syntactic, and contextual information whenever listening conditions permit (intelligible speech and meaningful sentences) in both read and conversational speech (White et al., 2012). Segmentation then simply "falls out" of word recognition, as implemented in TRACE. Sub-lexical cues intervene only

when lexical and contextual information is absent or incomplete – common in laboratory experiments but unusual in real life. And stress is used only as a last-resort, when intelligibility is too compromised to provide sufficient lexical, segmental, and subsegmental evidence.

With respect to ranking the weight of non-lexical cues, Newman and colleagues (2011) have argued that subsegmental cues (word-onset vowel glottalization) are as powerful as Norris and associates' (1997) Possible Word Constraint but that segmental probabilities (permissible syllable-final vowels) rank lower. In contrast, the robustness of distal prosody has been observed in the face of conflicting intensity, duration, *F*0 cues (Heffner et al., 2013), and even semantic information (Dilley et al., 2010), although there is evidence that these cues trade off as a function of their internal magnitude (Heffner et al., 2013).

Segmentation in the developing language system

The emergence of a mental lexicon is fundamental to how the speech signal is processed. Several decades of research have converged on the view that infants gradually integrate a range of cues in the service of segmentation. In isolation, these cues would be insufficient; together, they allow the child to segment in progressively greater detail. Several questions are relevant to understanding this process. First, to what extent is each cue reliably present in the infant's input? Second, can infants perceive and profitably use these cues? Finally, how do the cues interact? These questions are now addressed.

Cue presence

For a language user, the most elementary juncture marker is a silent pause. Silent pauses are usually found at boundaries between utterances, clauses, and phrases (Goldman-Eisler, 1972). Their contribution is particularly notable in infant-directed speech because utterances in that form are particularly short (Fernald et al., 1989; Snow, 1972). Likewise, corpus research has shown that words in isolation (e.g., "Milk!") constitute approximately 10% of the utterances that parents address to their infant (Brent & Siskind, 2001).

The features of infant-directed speech further increase its overall salience, providing a scaffold for infants to learn words by highlighting particular forms in the auditory stream. Infants' changing sensitivity reflects the emergence of a lexicon as well, with the development of stable word-form representations providing the necessary first step towards robust segmentation. For example, 6-month-old infants prefer the repetitive structure of infant-directed speech, whereas their earlier preference is for its prosodic elements (McRoberts et al., 2009). This shift in focus from prosodic to repetitive elements may be an indication of when infants transition from processing the general characteristics of speech to recognizing its components (i.e., words). This is consistent with findings showing that infants discriminate among words relatively early in life (Bergelson & Swingley, 2012; Tincoff & Jusczyk,

1999). Factors that add to the nascent lexicon are highly familiar items in the input (Bortfeld et al., 2005; Mandel et al., 1995), inclusion of word-like units in varying sentence frames (Gomez, 2002), and consistent production of the infant-directed rhythmic form (Ma et al., 2011).

Cue use

A pioneering demonstration of infant speech segmentation at 7.5 months (Jusczyk & Aslin, 1995) focused on the importance of emergent word knowledge well before those words are associated with meaning. In this study, infants were familiarized with a pair of novel words, such as "cup" and "feet", and then tested on their recognition of the words in passages. Results showed that infants listened longer to passages containing familiarized words compared with those containing novel words, demonstrating quite early segmentation ability. Subsequent studies using modifications of this paradigm have revealed limitations on infants' performance when words change in emotion, talker gender, or fundamental frequency (Bortfeld & Morgan, 2010; Houston & Jusczyk, 2000; Singh et al., 2004, 2008).

Other limitations include the finding that infants are able to segment only words that conform to the predominant stress pattern of their language (Jusczyk, 1999). For example, Weber and colleagues (2004) compared 4- and 5-month-old German-exposed infants with German-speaking adults specifically focusing on participants' mismatch negativity (MMN) responses to consonant-vowel-consonant-vowel sequences produced with either trochaic stress (stress placed on the initial syllable, which is predominant in German) or iambic stress (stress placed on the second syllable, which is atypical in German). Half of the participants experienced the trochaically stressed words as "standards" and the iambically stressed words as the MMN-dependent "deviants". The reverse was true for the other half of the participants. For the adults, an MMN response occurred whether the deviant was a trochaic or iambic sequence, suggesting that adults were sensitive to both stress patterns. However, for 5-month-olds, an MMN response was observed for deviant trochaic stimuli only, while neither stress type provoked a significant MMN response in the 4-month-olds. This suggests that by five months, infants are sensitive to the most common stress patterns of their exposure language, though they have yet to reach adult-like discrimination abilities for unfamiliar stress patterns (see also Hohle et al., 2009).

Cue interaction

Infants are sensitive to statistical regularities in their environment, using them as a guide to structure (e.g., Saffran et al., 1996b). However, statistical cues fail when variation in word length is introduced (Johnson & Tyler, 2010) or when they clash with subsegmental or stress cues (Johnson & Jusczyk, 2001). There are many other such interactions. For example, Mersad and Nazzi (2012) created an artificial language made of words that were varied (rather than fixed) in length. Consistent with

Johnson and Tyler's (2010) finding, 8-month-olds were hindered in their ability to segment words of varying lengths when presented with no other cues. However, they could segment them when the words were preceded with a familiar word (e.g., "maman"). In other words, infants were able to use the familiar item as a top-down guide for parsing the more complex signal.

With respect to the structure of cues used in the mature system, the developmental trajectory appears to follow a weighting system roughly opposite to that followed in adult processing (Figure 4.1). Stress and distributional regularities provide an early, albeit coarse first pass at the signal (Jusczyk et al., 1999; Saffran et al., 1996b). Although it is unclear whether sensitivity to stress spawns sensitivity to distributional regularities (Jusczyk, 1999) or vice versa (Thiessen & Saffran, 2003), these basic segmentation mechanisms appear to be progressively phased out by more subtle segmental and acoustic-phonetic cues by around 9 months of age (Morgan &

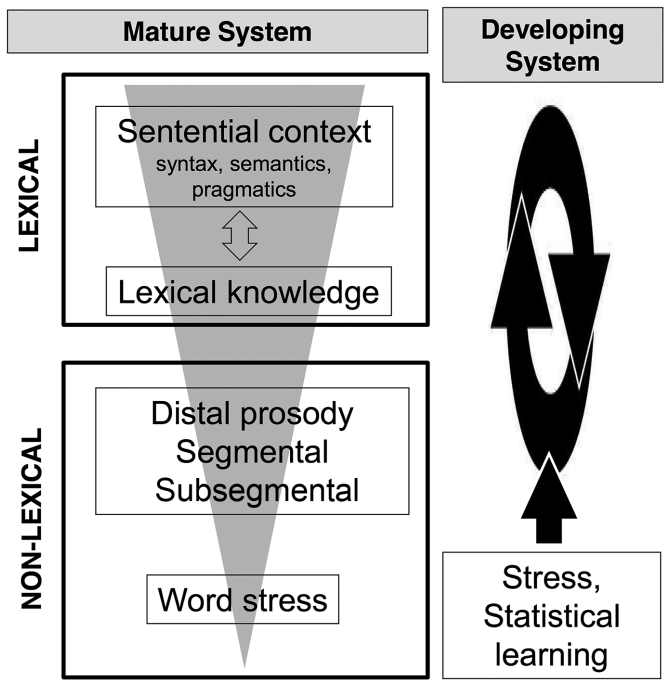

FIGURE 4.1 Schematic representation of primary segmentation cues in the mature and developing language systems. The relative weights of the cues in the mature speech system are symbolized by the width of the grey triangle (wider = greater weight). The developmental trajectory of the cues is illustrated by the black arrows. Reliance on stress and statistical learning cues is initially predominant. These, along with other non-lexical cues, promote the acquisition of a small vocabulary, which, in return, leads to the discovery of additional sub-lexical cues. Over time, reliance on non-lexical cues is downplayed in favour of lexical and contextual information (figure revised from Mattys et al. [2005]).

Saffran, 1995; Mattys et al., 1999). As these promote development of a lexicon, they become secondary to lexically driven segmentation, reflecting a move toward an increasingly adult-like weighting system.

These findings reveal both impressive capability and serious limitations on early word recognition. The second half of the first year of life clearly represents an important period of transition for language learning, as the ability to track spoken words in running speech and to detect repetitions of those words in different contexts is likely to be a prerequisite to later mapping of those words onto meaning. Indeed, longitudinal evidence for a link between early segmentation ability, on the one hand, and vocabulary growth at 2 years and higher-order language measures at 4 to 6 years on the other hand, confirms the constraining role of speech segmentation for successful language development (Newman et al., 2006). Important questions remain regarding the specific mechanisms linking the two and, perhaps more critically, how limitations in early segmentation might (negatively) impact the subsequent emergence and expansion of a vocabulary.

Segmentation in adverse (everyday) conditions

Not surprisingly, a majority of segmentation studies have focused on unrealistically ideal listening conditions. Comparatively few have examined how mechanisms are modulated by everyday conditions (noise, divided attention). Next we examine such modulations following Mattys and colleagues' (2012) typology of adverse conditions. A summary is depicted in Figure 4.2.

Source degradation

Perhaps the most detailed analysis of the link between source (speaker) degradation and speech segmentation comes from the work by Liss and colleagues on dysarthric speech perception. Dysarthrias are motor speech disorders subsequent to central or peripheral nervous system abnormalities such as Parkinson's disease or Huntington's chorea (for detailed nomenclature, see Darley et al., 1969), characterized by segmental, subsegmental, and suprasegmental disturbances. Liss et al. (1998) showed that healthy participants listening to English dysarthric speech made substantial use of the alternation between perceived weak and strong syllables to hypothesize word boundaries: perceived boundaries were inserted before strong syllables and deleted before weak syllables, in line with the dominant stress pattern of English (see also Cutler & Butterfield, 1992). However, Liss and associates (2000) found that stress-based segmentation was far more effective in dysarthrias that preserved rhythmic contrasts (hypokinetic) than dysarthrias that exhibited equal-and-even stress (ataxic). In a subsequent resynthesis study of non-dysarthric speech, Spitzer and colleagues (2007) demonstrated that neutralizing *F*0 (a suprasegmental cue) and blurring vowel identity (a segmental cue) had the most detrimental impact on listeners' adherence to stress-based segmentation. Equalizing vowel duration (a subsegmental and suprasegmental cue) had almost no effect. These analyses specify

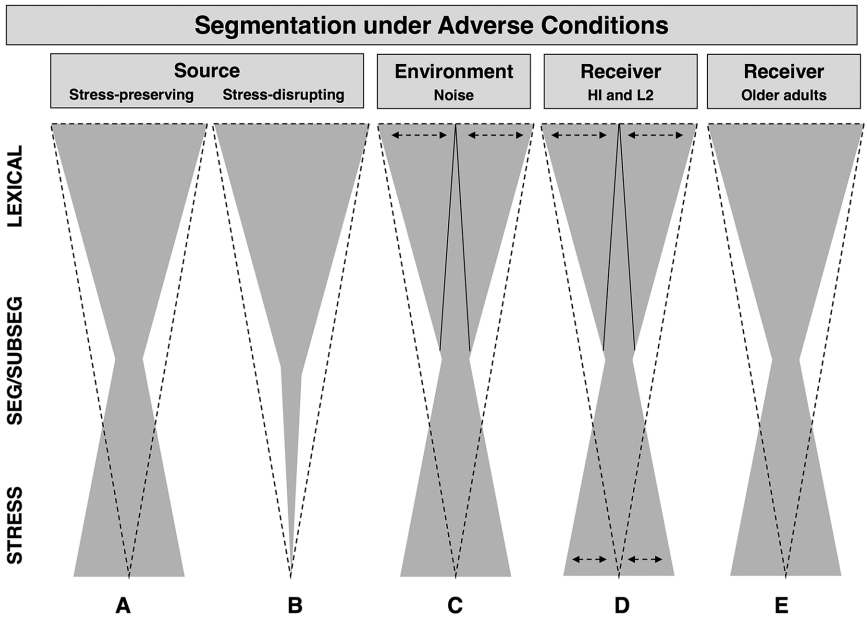

FIGURE 4.2 Schematic modulations of Figure 4.1 as a function of adverse listening conditions. As in Figure 4.1, the relative weights of the cues (lexical knowledge, segmental/subsegmental cues, stress) are symbolized by the width of the grey triangle (wider = greater weight). For comparison, the dark dashed outline illustrates optimal listening conditions, as shown in Figure 4.1. (A) Degradation at the source preserving the prosodic features necessary to perceive and use stress contrasts for segmentation, e.g., hypokinetic dysarthria or accented speech originating from the same rhythmic class as the target. (B) Degradation at the source disrupting or neutralizing stress contrast, e.g., ataxic dysarthria or accented speech originating from a different rhythmic class than the target. (C) Environmental degradation due to background noise. While reliance on stress is generally increased under noise, the contribution of lexical knowledge can be compromised in very low signal-to-noise ratios. (D) Receiver limitations due to hearing impairment (HI) or non-native (L2) knowledge of the target language. Reliance on lexical knowledge and stress in those cases is variable and primarily determined by the hearing impairment etiology and severity, the type of hearing device, the degree of overlap between L1 and the target language, and the level of L2 proficiency. (E) Older adults' generally increased reliance on lexical (and contextual) knowledge and stress contrast. For all depicted cases, note a reduction in reliance on segmental/subsegmental cues compared to segmentation in optimal conditions.

how naturally occurring speech degradations can turn lower-tier segmentation cues (Mattys et al., 2005) into powerful tools in the face of reduced intelligibility.

Accented speech constitutes another type of source degradation in that it departs from native or regional expectations on subsets (or all) of the non-lexical segmentation cues reviewed earlier – leaving aside the issue of lexical and syntactic deviations. The cost of processing a non-native/unfamiliar accent compared to

a native/familiar accent has been documented comprehensively (Munro & Derwing, 1995). Alongside phoneme miscategorization (Cutler et al., 2005), the most challenging cross-accent mismatch is likely to be prosodic. For instance, as speakers of syllable-timed languages (French) are likely to apply their native rhythm to the production of a non-native stress-timed language (English), word-initial stress cues normally used by native English listeners will no longer be available. Cases of misleading prosody can likewise be found between regional variations of a single language. For instance, Cutler (2012) raised the possibility that syllable-timed Singapore English (Deterding, 2001) might be more easily segmented by speakers of syllable-timed languages than by speakers of stress-timed languages. Thus, although on different scales, rhythmic structure violations due to motor-disordered speech and accent have a lot in common in terms of the reweighting of segmentation strategies that listeners apply to cope with the deviations.

Environmental degradation

This category includes degradations of speech due to competing signals (noise), acoustic distortions caused by the physical environment (reverberation), and impoverishment by the transmission medium (spectral filtering through a telephone). Segmentation research has focused mainly on broadband noise. Mattys and associates (2005) found that the effectiveness of lexical-semantic knowledge drops steadily as a function of noise level, probably reflecting the increasingly diffuse lexical activity that results from inaccurately encoded sensory information. The greater reliance on non-lexical cues resulting from lexical-access failure depends on the regions of the frequency spectrum where signal and noise overlap. Generally speaking, non-lexical cues consisting of broad variations over relatively long stretches of the signal fare best. For example, suprasegmental cues such as stress and *F*0 movements are resilient to 5- to 10-dB signal-to-noise ratios (Mattys, 2004; Smith et al., 1989; Welby, 2007). In contrast, coarticulatory cues and transitional probabilities show greater vulnerability (Mattys et al., 2005), with the former more fragile than the latter (Fernandes et al., 2007).

The generalizability of these results is unclear, however. Disruption caused by even the loudest broadband steady-state noise is negligible if it coincides with energy dips or highly redundant portions of the signal. Conversely, a brief burst of energy in a critical frequency range could have consequences cascading from misheard allophones to sentence misinterpretation. Thus, unlike cue re-ranking in the face of dysarthric speech, cue re-ranking under noise is highly dependent on complex interactions between the spectro-temporal structure of the noise and the phonetic, lexical, and semantic composition of the signal (Bronkhorst, 2000).

It should be noted that correctly perceiving speech in the presence of noise should be particularly challenging for infants and young children. And indeed, 66% of parents report talking to their infants while other family members are talking at the same time (Barker & Newman, 2004). Not surprisingly, signal-to-noise ratio (SNR) is an important factor to consider. Children presumably cannot rely on

top-down information to the same extent as adults (Newman, 2006; Nittrouer & Boothroyd, 1990), which means that they should weigh other cues in the signal more heavily (Fallon et al., 2000). This, in turn, means that they may be affected by noise much more or at least differently than adults. As evidence of this, Newman (2004) found that 5-month-old infants failed to recognize their own names in a multi-talker stream unless the masking signal was at least 10 dB below the target signal. Moreover, with a 0 dB SNR, she found that infants performed at chance and continued to do so until at least the end of the first year. But this changes quickly as children begin to acquire the lexicon. For example, children's word learning between 2.5 and 3 years was equally likely whether the learning took place in quiet, +5, or 0 dB SNR (Dombroski & Newman, 2014). Although these latter findings bode well for language development in spite of the varied environments in which children find themselves, the input they receive is a critical factor in the rate and success of this development.

Receiver limitations

In this section, we review findings from hearing-impaired, non-native, and older listeners, groups with contrasting, though not fully orthogonal segmentation challenges.

Segmentation in hearing-impaired listeners

Relatively little is known about the segmentation limitations faced by hearing aid and cochlear implant (CI) users. Research suggests that hearing device users rely on higher-order knowledge to fill gaps in the impoverished signal. For instance, using noise-vocoded speech (i.e., substitution of frequency bands with amplitude-modulated noise, Shannon et al., 1995) as a simulation of CI input, Davis and colleagues (2005) showed that low-level interpretation of the distorted signal was better when it contained known and meaningful words than when it contained non-words. Thus, to the extent that contextual information is available from the distorted signal, it is likely that lexically driven segmentation mechanisms are primary.

Because hearing devices alter spectral rather than temporal aspects of the signal, segmental, allophonic, and *F*0 cues to word boundaries are more affected than durational cues. Given the contribution of *F*0 to the perception of stress contrast in English (Spitzer et al., 2007), stress-based segmentation could potentially be seriously compromised by CI. Indeed, Spitzer et al. (2009) showed that the detrimental effect of flattening *F*0 on stress-based segmentation was more pronounced for normal-hearing and hearing-impaired individuals with residual low-frequency hearing than for CI users. The authors interpreted this result as showing that *F*0 representation in the CI group was poor to start with, and hence flattening *F*0 had little effect. This conclusion, however, must be restricted to languages in which stress is realised mainly through *F*0. Stress-based segmentation of languages where stress correlates with syllable lengthening rather than *F*0 is likely to remain comparatively robust.

To a large extent, the segmentation problems that CI users face are comparable, albeit on a more dramatic scale, to those faced by people listening to accented speech: The distortions affect some sounds more than others, and they do so systematically. The predictable nature of such distortions makes correction through perceptual learning possible, unlike conditions in which the signal is masked by random noise. Thus, the study of speech segmentation through hearing devices is more likely to make headway if it is investigated in the context of progressive cue recalibration than if it is seen solely from the perspective of signal impoverishment.

Segmentation in non-native (L2) listeners

Although comparable to the problems faced by infant learners, the segmentation challenge faced by L2 listeners involves substantial differences. On the one hand, L2 listeners know a lot more about language (in general) than infants do, and they are likely to be literate. On the other hand, the segmentation cues of a native language (L1) can be at odds with those of L2, thus making the mastery of L2 a matter of both learning the segmentation cues of L2 and suppressing those of L1.

Not unlike infants, non-native speakers use their growing L2 lexicon to discover new words. For instance, White and associates (2010) showed that adult L1 Hungarian speakers learning English used known English words to segment adjacent sound sequences. Importantly, this segmentation-by-lexical-subtraction process applied even when stress cues were inconsistent with it. Given that Hungarian is a fixed-stress language (initial stress in all content words), this provides clear evidence for the primacy of lexically driven (over non-lexical) cues.

The success of L2 segmentation in the absence of lexical information depends on the respective hierarchies of non-lexical cues in L1 and L2. In general, languages with similar junctural regularities (comparable phonotactic rules or stress placement) will facilitate the transfer of segmentation strategies, whereas those with divergent regularities may create interference. The cost of persevering with L2-inappropriate cues has been shown for phonotactic regularities (Weber & Cutler, 2006), acoustic-phonetic cues (Tremblay & Spinelli, 2014) and stress (Sanders et al., 2002a; Tremblay et al., 2012). Thus, the process of becoming a proficient L2 listener involves not only vocabulary growth and syntactic learning (Sanders et al., 2002a, 2003) but also increased control over potentially contradictory sets of non-lexical segmentation strategies.

Segmentation in older listeners

There are two theoretical reasons to be interested in segmentation in older listeners. First, aging is accompanied with a steady decline in auditory abilities, particularly high-frequency perception, temporal processing, and sound stream segregation (CHABA, 1988). Second, vocabulary shows comparatively little decline (Singer et al., 2003). This combination of decline and stability suggests that segmentation is likely to become increasingly lexically driven in later years, with segmental and

subsegmental cues becoming less impactful. Interestingly, the perception of suprasegmental patterns such as stress is relatively well preserved (Baum, 2003; Wingfield et al., 2000), thus making segmentation in older adults possibly a lexico-prosodic matter.

Finally, the large literature on perceptual and cognitive decline in older adults indicates a heightened sensitivity to noise and cognitive load (Burke & Shafto, 2008). Therefore, segmentation challenges are likely to be amplified under conditions of environmental degradation, especially when these tax working memory (Rönnberg et al., 2010), another declining faculty in older adults (McDaniel et al., 2008). Recent research by Weiss and colleagues (2010) suggests that resolving conflicts between segmentation cues might require increased processing resources. Given age-related decline in cognitive capacity (Craik & Byrd, 1982), cue conflict is likely to be particularly challenging for older adults.

Conclusion

Speech segmentation is an elusive construct insofar as it is often little more than a by-product of word recognition. In that sense, segmentation is implicitly built into models of speech recognition. However, segmentation arises as a concrete challenge whenever the interface between signal and lexical representations fails. In a narrow sense, the goal of this chapter was to review the factors at the root of such failures. These included lexical embeddedness, a developing lexicon or language system (L1 or L2), input degradation, and auditory limitations. Our contention is that the extent to which segmental, subsegmental, and suprasegmental cues influence segmentation can be predicted based on the specific characteristics of each of these factors, as well as the statistical structure of the language. Modelling speech segmentation remains an ambitious enterprise, however, because the relative weights of lexical and non-lexical information are likely to change on a moment-by-moment basis during the unfolding of speech. Online experimental measures of segmentation such as eye movements (e.g., Tremblay, 2011) and electrophysiological brain responses (e.g., Cunillera et al., 2009; Sanders et al., 2002b) are providing finer analyses of such temporal fluctuations. Likewise, significant advances in understanding the segmentation process can be made through a better appreciation of the dynamic flow of information between layers of the language system. Recent studies relying on causality analyses of magnetoencephalographic (MEG) and electroencephalographic (EEG) data during speech perception tasks (Gow et al., 2008, 2009) are providing a useful hypothetical functional architecture in which segmentation can be further analysed.

References

Banel, M. H., & Bacri, N. (1997). Reconnaissance de la parole et indices de segmentation métriques et phonotactiques. *L'Année psychologique*, *97*, 77–112.

Barker, B. A., & Newman, R. S. (2004). Listen to your mother! The role of talker familiarity in infant streaming. *Cognition*, *94*, B45–53.

Baum, S. R. (2003). Age differences in the influence of metrical structure on phonetic identification. *Speech Communication*, *39*, 231–242.

Bergelson, E., & Swingley, D. (2012). At 6–9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Sciences*, *109*, 3253–3258.

Bortfeld, H., & Morgan, J. (2010). Is early word-form processing stress-full? How natural variability supports recognition. *Cognitive Psychology*, *60*, 241–266.

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

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

Bronkhorst, A. W. (2000). The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions. *Acta Acustica*, *86*, 117–128.

Brown, M., Salverda, A. P., Dilley, L. C., & Tanenhaus, M. K. (2011). Expectations from preceding prosody influence segmentation in online sentence processing. *Psychonomic Bulletin & Review*, *18*, 1189–1196.

Burke, D. M., & Shafto, M. A. (2008). Language and aging. In F. I. M. Craik & T. A. Salthouse (Eds.), *The handbook of aging and cognition* (3rd ed., pp. 373–443). New York: Psychology Press.

Cairns, P., Shillcock, R., Chater, N., & Levy, J. (1997). Bootstrapping word boundaries: A bottom-up corpus-based approach to speech segmentation. *Cognitive Psychology*, *33*, 111–153.

Christophe, A., Peperkamp, S., Pallier, C., Block, E., & Mehler, J. (2004). Phonological phrase boundaries constrain lexical access I. Adult data. *Journal of Memory and Language*, *51*, 523–547.

Committee on Hearing and Bioacoustics and Biomechanics (CHABA). (1988). Speech understanding and aging. *Journal of the Acoustical Society of America*, *83*, 859–895.

Craik, F. I. M., & Byrd, M. (1982). Aging and cognitive deficits: The role of attentional resources. In F. I. M. Craik & S. Trehub (Eds.), *Aging and cognitive processes* (pp. 191–211). New York: Plenum Press.

Cunillera, T., Càmara, E., Toro, J. M., Marco-Pallares, J., Sebastián-Galles, N., Ortiz, H., Pujol, J., & Rodríguez-Fornells, A. (2009). Time course and functional neuroanatomy of speech segmentation in adults. *Neuroimage*, *48*, 541–553.

Cutler, A. (2012). *Native listening: Language experience and the recognition of spoken words*. Cambridge, MA: MIT Press.

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

Cutler, A., & Norris, D. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance*, *14*, 113–121.

Cutler, A., Smits, R., & Cooper, N. (2005). Vowel perception: Effects of non-native language vs. non-native dialect. *Speech communication*, *47*, 32–42.

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*, 165–185.

Darley, F. L., Aronson, A. E., & Brown, J. R. (1969). Differential diagnostic patterns of dysarthria. *Journal of Speech, Language, and Hearing Research*, *12*, 246–269.

Davis, M. H., Johnsrude, I. S., Hervais-Adelman, A., Taylor, K., & McGettigan, C. (2005). Lexical information drives perceptual learning of distorted speech: Evidence from the comprehension of noise-vocoded sentences. *Journal of Experimental Psychology: General*, *134*, 222.

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*, 218–244.

Deterding, D. (2001). The measurement of rhythm: A comparison of Singapore and British English. *Journal of Phonetics*, *29*, 217–230.

Dilley, L. C., Mattys, S. L., & Vinke, L. (2010). Potent prosody: Comparing the effects of distal prosody, proximal prosody, and semantic context on word segmentation. *Journal of Memory and Language*, *63*(3), 274–294.

Dilley, L. C., & McAuley, J. D. (2008). Distal prosodic context affects word segmentation and lexical processing. *Journal of Memory and Language*, *59*, 294–311.

Dombroski, J., & Newman, R. (2014). Toddlers' ability to map the meaning of new words in multi-talker environments. *Journal of the Acoustical Society of America*, *136*, 2807–2815.

Fallon, M., Trehub, S. E., & Schneider, B. A. (2000). Children's perception of speech in multitalker babble. *Journal of the Acoustical Society of America*, *108*, 3023–3029.

Fernald, A., Taeschner, T., Dunn, J., Papousek, M., 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*, 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*, 856–864.

Frauenfelder, U. H., & Peeters, G. (1990). Lexical segmentation in TRACE: An exercise in simulation. In G. Altmann (Ed.), *Cognitive models of speech processing: Psycholinguistic and computational perspectives* (pp. 50–86). Cambridge, MA: MIT Press.

Goldman-Eisler, F. (1972). Pauses, clauses, sentences. *Language and Speech*, *15*, 103–113.

Goldwater, S., Griffiths, T. L., & Johnson, M. (2009). A Bayesian framework for word segmentation: Exploring the effects of context. *Cognition*, *112*, 21–54.

Gomez, R. (2002). Variability and detection of invariant structure. *Psychological Science*, *13*, 431–436.

Gow Jr, D. W., Keller, C. J., Eskandar, E., Meng, N., & Cash, S. S. (2009). Parallel versus serial processing dependencies in the perisylvian speech network: A Granger analysis of intracranial EEG data. *Brain and Language*, *110*, 43–48.

Gow Jr, D. W., Segawa, J. A., Ahlfors, S. P., & Lin, F. H. (2008). Lexical influences on speech perception: A Granger causality analysis of MEG and EEG source estimates. *Neuroimage*, *43*, 614–623.

Heffner, C. C., Dilley, L. C., McAuley, J. D., & Pitt, M. A. (2013). When cues combine: How distal and proximal acoustic cues are integrated in word segmentation. *Language and Cognitive Processes*, *28*, 1275–1302.

Hohle, B., Bijeljac-Babic, R., Herold, B., Weissenborn, J., & Nazzi, T. (2009). Language specific prosodic preferences during the first half year of life: Evidence from German and French infants. *Infant Behavior and Development*, *32*, 262–274.

Houston, D. M., & Jusczyk, P. W. (2000). The role of talker specific information in word segmentation by infants. *Journal of Experimental Psychology: Human Perception and Performance*, *26*, 1570–1582.

Janse, E., Nooteboom, S. G., & Quené, H. (2007). Coping with gradient forms of /t/-deletion and lexical ambiguity in spoken word recognition. *Language and Cognitive Processes*, *22*, 161–200.

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*, 548–567.

Johnson, E. K., & Tyler, M. D. (2010). Testing the limits of statistical learning for word segmentation. *Developmental Science*, *13*, 339–345.

Jusczyk, P. W. (1999). How infants begin to extract words from speech. *Trends in Cognitive Sciences*, *3*, 323–328.

Jusczyk, P. W., & Aslin, R. N. (1995). Infants' detection of the sound patterns of words in fluent speech. *Cognitive Psychology*, *29*, 1–23.

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

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*, 509–529.

Liss, J. M., Spitzer, S., Caviness, J. N., Adler, C., & Edwards, B. (1998). Syllabic strength and lexical boundary decisions in the perception of hypokinetic dysarthric speech. *The Journal of the Acoustical Society of America*, *104*, 2457–2466.

Liss, J. M., Spitzer, S. M., Caviness, J. N., Adler, C., & Edwards, B. W. (2000). Lexical boundary error analysis in hypokinetic and ataxic dysarthria. *The Journal of the Acoustical Society of America*, *107*, 3415–3424.

Ma, W., Golinkoff, R., Houston, D. M., & Hirsh-Pasek, K. (2011). Word learning in infant- and adult-directed speech. *Language Learning and Development*, 7, 185–201.

Mandel, D. R., Jusczyk, P. W., & Pisoni, D. B. (1995). Infants' recognition of the sound patterns of their own names. *Psychological Science*, *6*, 314–317.

Mattys, S. L. (2004). Stress versus coarticulation: Towards an integrated approach to explicit speech segmentation. *Journal of Experimental Psychology: Human Perception and Performance*, *30*, 397–408.

Mattys, S. L., Davis, M. H., Bradlow, A. R., & Scott, S. K. (2012). Speech recognition in adverse conditions: A review. *Language and Cognitive Processes*, *27*, 953–978.

Mattys, S. L., Jusczyk, P. W., Luce, P. A., & Morgan, J. L. (1999). Phonotactic and prosodic effects on word segmentation in infants. *Cognitive Psychology*, *38*, 465–494.

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*, 960–977.

Mattys, S. L., White, L., & Melhorn, J. F. (2005). Integration of multiple segmentation cues: A hierarchical framework. *Journal of Experimental Psychology: General*, *134*, 477–500.

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

McDaniel, M., Einstein, G., & Jacoby, L. (2008). New considerations in aging and memory: The glass may be half full. In F. I. M. Craik & T. A. Salthouse (Eds.), *The handbook of aging and cognition* (3rd ed., pp. 251–310). New York: Psychology Press.

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

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*, 309–331.

McRoberts, G. W., McDonough, C., & Lakusta, L. (2009). The role of verbal repetition in the development of infant speech preferences from 4 to 14 months of age. *Infancy*, *14*, 162–194.

Mersad, K., & Nazzi, T. (2012). When mommy comes to the rescue of statistics: Infants combine top-down and bottom-up cues to segment speech. *Language Learning and Development*, *8*, 303–315.

Monaghan, P., White, L., & Merkx, M. M. (2013). Disambiguating durational cues for speech segmentation. *Journal of the Acoustical Society of America*, *134*, EL45–EL51.

Morgan, J. L., & Saffran, J. R. (1995). Emerging integration of sequential and suprasegmental information in preverbal speech segmentation. *Child Development*, *66*, 911–936.

Munro, M. J., & Derwing, T. M. (1995). Processing time, accent, and comprehensibility in the perception of native and foreign-accented speech. *Language and Speech*, *38*, 289–306.

Newman, R. S. (2004). Perceptual restoration in children versus adults. *Applied Psycholinguistics*, *25*, 481–493.

Newman, R. S. (2006). Perceptual restoration in toddlers. *Perception & Psychophysics*, *68*, 625–642.

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*, 643–655.

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*, 460–476.

Nittrouer, S., & Boothroyd, A. (1990). Context effects in phoneme and word recognition by young children and older adults. *Journal of the Acoustical Society of America*, *87*, 2705–2715.

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

Norris, D., & McQueen, J. M. (2008). Shortlist B: A Bayesian model of continuous speech recognition. *Psychological Review*, *115*, 357–395.

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

Ohala, J. J. (1990). The phonetics and phonology of aspects of assimilation. *Papers in Laboratory Phonology*, *1*, 258–275.

Otake, T., Hatano, G., Cutler, A., & Mehler, J. (1993). Mora or syllable? Speech segmentation in Japanese. *Journal of Memory and Language*, *32*, 258–278.

Roach, P., Knowles, G., Varadi, T., & Arnfield, S. (1993). Marsec: A machine-readable spoken English corpus. *Journal of the International Phonetic Association*, *23*, 47–54.

Rönnberg, J., Rudner, M., Lunner, T., & Zekveld, A. A. (2010). When cognition kicks in: Working memory and speech understanding in noise. *Noise and Health*, *12*, 263–269.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996b). Statistical learning by 8-month-old infants. *Science*, *274*, 1926–1928.

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

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

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

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

Schuppler, B., Ernestus, M., Scharenborg, O., & Boves, L. (2011). Acoustic reduction in conversational Dutch: A quantitative analysis based on automatically generated segmental transcriptions. *Journal of Phonetics*, *39*, 96–109.

Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science*, *270*, 303–304.

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

Singer, T., Verhaeghen, P., Ghisletta, P., Lindenberger, U., & Baltes, P. B. (2003). The fate of cognition in very old age: Six-year longitudinal findings in the Berlin Aging Study (BASE). *Psychology and Aging*, *18*, 318–331.

Singh, L., Morgan, J. L., & White, K. S. (2004). Preference and processing: The role of speech affect in early spoken word recognition. *Journal of Memory and Language*, *51*, 173–189.

Singh, L., Nestor, S. S., & Bortfeld, H. (2008). Overcoming effects of variation on infant word recognition: Influences on word familiarity. *Infancy*, *13*, 57–74.

Smith, M. R., Cutler, A., Butterfield, S., & Nimmo-Smith, I. (1989). The perception of rhythm and word boundaries in noise-masked speech. *Journal of Speech, Language, and Hearing Research*, *32*, 912–920.

Snow, C. E. (1972). Mothers' speech to children learning language. *Child Development*, *43*, 549–565.

Spinelli, E., McQueen, J. M., & Cutler, A. (2003). Processing resyllabified words in French. *Journal of Memory and Language*, *48*, 233–254.

Spitzer, S. M., Liss, J. M., & Mattys, S. L. (2007). Acoustic cues to lexical segmentation: A study of resynthesized speech. *Journal of the Acoustical Society of America*, *122*, 3678–3687.

Spitzer, S., Liss, J., Spahr, T., Dorman, M., & Lansford, K. (2009). The use of fundamental frequency for lexical segmentation in listeners with cochlear implants. *The Journal of the Acoustical Society of America*, *125*, EL236–EL241.

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

Tagliapietra, L., Fanari, R., De Candia, C., & Tabossi, P. (2009). Phonotactic regularities in the segmentation of spoken Italian. *The Quarterly Journal of Experimental Psychology*, *62*, 392–415.

Thiessen, E. D., & Saffran, J. R. (2003). When cues collide: Use of stress and statistical cues to word boundaries by 7- to 9-month-old infants. *Developmental Psychology*, *39*, 706–716.

Tincoff, R., & Jusczyk, P. W. (1999). Some beginnings of word comprehension in 6-month-olds. *Psychological Science*, *10*, 172–175.

Toro, J. M., Sebastián-Gallés, N., & Mattys, S. L. (2009). The role of perceptual salience during the segmentation of connected speech. *European Journal of Cognitive Psychology*, *21*, 786–800.

Tremblay, A. (2011). Learning to parse liaison-initial words: An eye-tracking study. *Bilingualism: Language and Cognition*, *14*, 257–279.

Tremblay, A., Coughlin, C. E., Bahler, C., & Gaillard, S. (2012). Differential contribution of prosodic cues in the native and non-native segmentation of French speech. *Laboratory Phonology*, *3*, 385–423.

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*, 310–337.

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

Vaissière, J. (1983). Language-independent prosodic features. In A. Cutler & D. L. Ladd (Eds.), *Prosody: Models and measurements* (pp. 53–66). Berlin/Heidelberg: Springer.

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.

Vroomen, J., Van Zon, M., & de Gelder, B. (1996). Cues to speech segmentation: Evidence from juncture misperceptions and word spotting. *Memory & Cognition*, *24*, 744–755.

Warner, N., & Weber, A. (2001). Perception of epenthetic stops. *Journal of Phonetics*, *29*, 53–87.

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

Weber, C., Hahne, A., Friedrich, M., & Friederici, A. D. (2004). Discrimination of word stress in early infant perception: Electrophysiological evidence. *Cognitive Brain Research*, *18*, 149–161.

Weiss, D. J., Gerfen, C., & Mitchel, A. D. (2010). Colliding cues in word segmentation: The role of cue strength and general cognitive processes. *Language and Cognitive Processes*, *25*, 402–422.

Welby, P. (2007). The role of early fundamental frequency rises and elbows in French word segmentation. *Speech Communication*, *49*, 28–48.

White, L. (2014). Communicative function and prosodic form in speech timing. *Speech Communication*, *63*, 38–54.

White, L., Mattys, S. L., & Wiget, L. (2012). Segmentation cues in conversational speech: Robust semantics and fragile phonotactics. *Frontiers*, *3*, 1–9.

White, L., Melhorn, J. F., & Mattys, S. L. (2010). Segmentation by lexical subtraction in Hungarian L2 speakers of English. *Quarterly Journal of Experimental Psychology*, *63*, 544–554.

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*, 1707–1717.

Wingfield, A., Lindfield, K. C., & Goodglass, H. (2000). Effects of age and hearing sensitivity on the use of prosodic information in spoken word recognition. *Journal of Speech, Language, and Hearing Research*, *43*, 915–925.