9 Word Stress in Speech Perception

ANNE CUTLER[1] AND ALEXANDRA JESSE[2]

[1]Western Sydney University, Australia
[2]University of Massachusetts Amherst, United States

Stress denotes greater relative salience of some linguistic elements compared with others, within larger units of speech. The term can be applied to different speech domains. Thus some words are stressed more than others within a sentence, and some syllables are stressed more than others within words. What controls the positioning of stress is not random; at the sentence level, it is largely determined by the relative importance of sentence components for the ongoing discourse (information structure), and at the lexical level it is fully determined by word phonology. While information structure can reasonably be held to originate outside the specifically linguistic domain and hence to have claim to universality, word phonology is highly language specific, and not all languages have word-level stress.

In languages that do have word stress, segmentally matched stressed and unstressed syllables (such as the first syllables of English *camper* and *campaign*) generally differ in multiple acoustic dimensions. However, the consequences of stress placement for speech perception are not simply a function of these acoustic variations. The next section of this chapter, "Lexical stress and the vocabulary," shows that the phonology of the language in question, and the resulting vocabulary structure, determine what use is made of stress-related acoustic information. If word stress varies, cues to stress location can help identify spoken words and can modulate the activation and competition processes involved in this; but as demonstrated in the following section, "Spoken-word identification," even related and in principle quite similar languages can vary greatly in how such cues are exploited, with the underlying driver of cue use being, indeed, the language-specific vocabulary patterns. In "New horizons for stress in speech perception," the currently very active state of the field of word-stress perception research is illustrated by innovative data from multisensory perception and from perception

The Handbook of Speech Perception, Second Edition.
Edited by Jennifer S. Pardo, Lynne C. Nygaard, Robert E. Remez, and David B. Pisoni.

of degraded speech. The chapter concludes with a summary and the prediction that this extensive activity, spanning not only new techniques but also many language groups, will produce substantial and detailed new knowledge on the role of word stress in speech perception.

Among the languages with word-level stress, some allow stress placement to vary within the word while others do not. Languages where stress placement in words can vary are said to have "lexical stress." These languages in principle allow the relative stress level of syllables to distinguish otherwise (i.e. segmentally) identical word forms. Languages where stress placement cannot vary within the word ("fixed-stress" languages) obviously preclude such a lexically contrastive function for stress. In the latter class of languages, note that, while stress always falls in the same place in the word, that place itself is language-specifically defined: the initial syllable in Finnish, the final syllable in Turkish, the antepenultimate syllable in Macedonian, and so on. Thus there is no universal pattern either for the appearance of word stress (some languages don't have it at all), or for its realization when it does appear (it can be fixed or it can vary), and, importantly, there can therefore be no universal pattern for its role in speech perception (only in some languages can it be contrastive). The story that this chapter has to tell, in other words, is at its core one of language specificity.

Lexical stress is the variety of stress that English has, like its West Germanic language relatives; the kind of minimal pairs this allows include noun–verb distinctions such as *PERvert* (noun; capital letters indicate primary stress) versus *perVERT* (verb), but also word pairs unrelated in meaning, such as *FOREgoing* and *forGOing*. In all such cases, the segments are all the same and only the stress placement differs. In lexical-stress languages, the stress pattern of every polysyllabic word is lexically determined, that is, is part of the phonological representation of how speakers ought to produce the word. In fact, minimal whole-word pairs are not numerous in any lexical-stress language. Far more commonly, speech perception involves minimally paired initial syllables that differ in stress, such as the first syllables of the English word *CAMper* versus *camPAIGN*. Such segmentally identical syllables, one stressed and the other not, differ in how they are uttered.

The difference in lexical stress is realized in several acoustic dimensions, as can be seen in Figure 9.1. This figure shows waveforms and spectrograms for a male speaker of American English saying the *pervert* pair in the context *Say the word pervert again*. The top three panels show the verb reading *perVERT*, the lower three the noun reading *PERvert*. Although the syllables have the same segmental structure in each member of the pair, including full vowels in each syllable, the acoustic realization can be seen to be clearly different in the three suprasegmental dimensions duration, intensity, and fundamental frequency (F0). For each syllable, the stressed version is longer and louder (most immediately visible in the waveform), and presents more F0 movement (see the spectrograms).

In the early days of systematic research on speech perception (when most phonetic research was done in countries with a West Germanic language), stress perception was a topic, with the consensus opinion converging on the conclusion that all of the above cues were used, separately or in combination. Searches for a

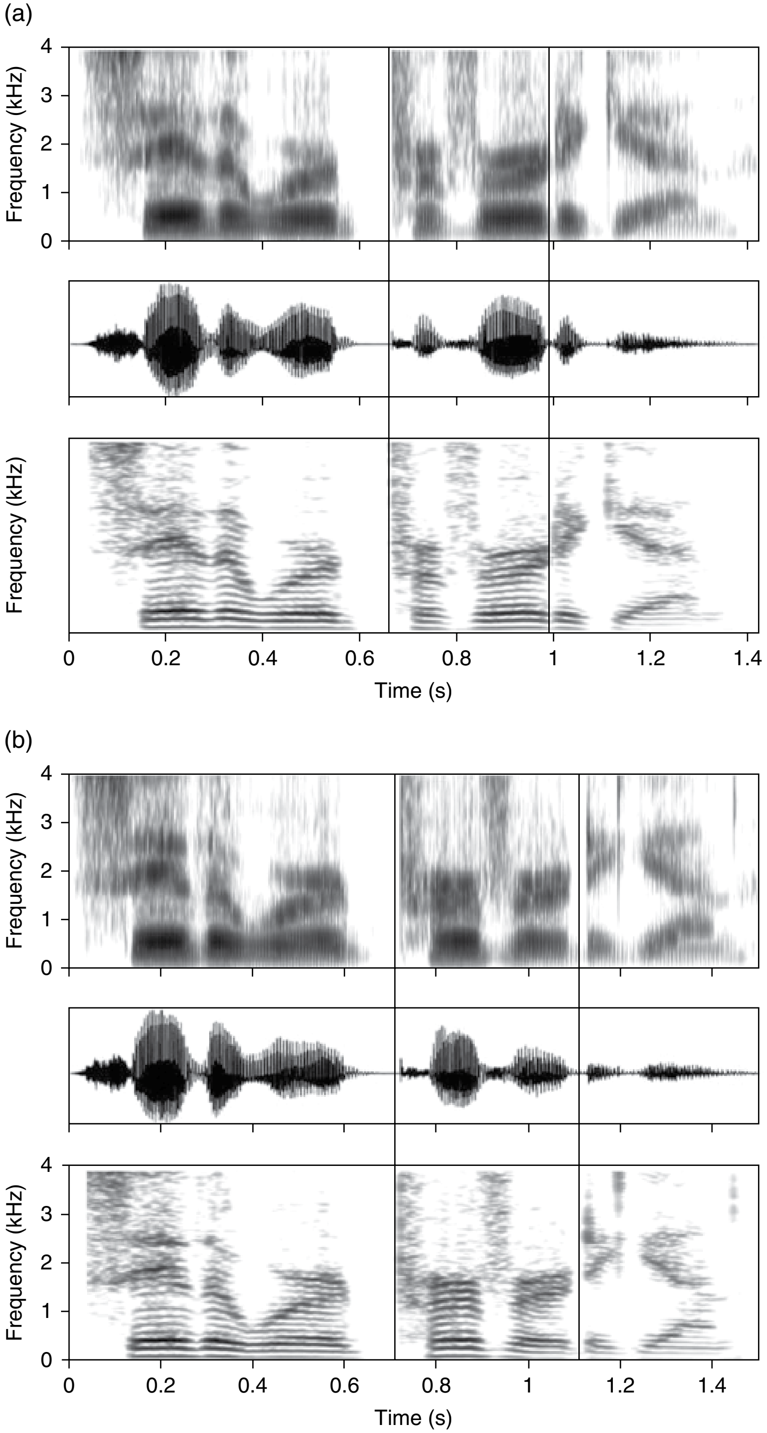

Figure 9.1 Sound spectrograms of the words *perVERT* (a) and *PERvert* (b) in the carrier sentence *Say the word . . . again*, recorded by a male speaker of American English. Each figure consists of three display panels: (top) a broad-band spectrogram; (middle) a waveform display; and (bottom) a narrow-band spectrogram. Vertical lines indicate onset and offset of *pervert*. The figure is modeled on a figure first presented by Lehiste and Peterson (1959, p. 434).

single unifying factor in the perception of stress, such as articulatory or perceptual effort, were unsuccessful (more detail of this early work can be found in Cutler, 2005). Extending the stress investigations to other languages was critical in showing that the phonological system of a language determines how stress is realized and hence perceived. For example, in some tone languages F0 may not distinguish stress as it is reserved to convey tone (see Potisuk, Gandour, & Harper, 1996, for Thai), and in languages with vowel quantity distinctions, duration may be likewise reserved to convey vowel identity and hence be unavailable as a stress cue (see Berinstein, 1979, for Mayan languages). Note also that not only the suprasegmental dimensions realize lexical stress, as segmental structure can also vary systematically with stress; in English, for instance, any vowel can be either stressed or not except for the central vowel schwa which is necessarily unstressed. Again, this is dependent on language-specific phonology, as lexical-stress languages without such a vowel in their phonemic repertoire will realize stress only suprasegmentally (e.g. Spanish), while in languages with schwa that vowel may be associated with unstressed syllables only (as in English or German), or may be a stressable vowel like any other (as in Welsh; Williams, 1985).

In fixed-stress languages, there is sometimes little measurable acoustic difference between syllables in the stressed position and segmentally identical syllables that occur in another position in the word (Suomi, Toivanen, & Ylitalo, 2003, for Finnish; Dogil, 1999, for Polish, which has fixed penultimate stress). This does not exclude a speech-perception function for fixed stress, however; obvious options remain. Where the designated stress placement is at a word's edge, for instance, it may appear to be useful for identifying word boundaries in running speech. Since that would require syllables bearing fixed stress to be reliably distinguished from those without stress (i.e. the stress to be realized acoustically in a consistent manner), the acoustic evidence argues against it. The designated "stressed" syllables are, however, the location for sentence-level accents that fall upon the word in question, and when these sentence-level accents are realized, the acoustic effects that this brings about will be informative in sentence-level processing. It is for word-level processing that fixed stress offers little support.

In fact, when speakers of fixed-stress languages are asked to perform tasks involving word-level stress perception, they exhibit in many cases what has been called "stress deafness" (although the term is inaccurate; see the section on "Spoken-word identification"). As was first demonstrated by Dupoux et al. (1997) for French, in which clitic phrases are accented on a fixed location (the final syllable), such listeners often err on decisions about stress placement in heard nonsense material. For instance, they have high error rates in an ABX task in which a third token must be matched to one of two preceding tokens (e.g. *bopeLO boPElo bopeLO*). Further work by Dupoux and colleagues (Dupoux & Peperkamp, 2002; Dupoux, Peperkamp, & Sebastián-Gallés, 2001; Peperkamp, Vendelin, & Dupoux, 2010) showed the same contrasts to be predictably easy for speakers of Spanish (a lexical-stress language) but just as hard as in French for speakers of the fixed-stress language Finnish, and almost as hard for speakers of Polish and Hungarian, also with fixed stress. Dupoux and Peperkamp accounted for the gradations of difference across the fixed-stress groups in terms of language-specific phonology and its learnability (Polish and Hungarian accentual rules affect different word classes in different ways, so their stress systems will be harder to learn than the French or Finnish systems, and this greater learning challenge leaves a trace of stress cue sensitivity in adulthood, detectable in the perception experiments).

It is clear that language specificity characterizes speech perception at the word level. What is involved in recognizing spoken words in everyday conversations is twofold: segmentation (identifying where each word begins, given that speech is continuous and word boundaries are not reliably marked), and lexical access (identifying the word from all the other word forms it might be, including of course identifying where it ends). For stress to play a role, acoustic cues must in any case be present. If that is the case, then segmentation can involve stress in its demarcative role (Cutler & Carter, 1987; Tyler & Cutler, 2009). Identification consists, critically, in the rejection of alternative word candidates. Recognition of a spoken word therefore depends to a significant degree on how many other words are like it, in particular its onset overlap competitors (e.g. *camp* and *campanology* for *campaign*) but also its embeddings (*am* and *pain* in *campaign*). The term "competitor" here suggests that candidate words compatible with some portion of the incoming speech signal actively compete with one another, and its use in this context has arisen from findings showing inhibitory effects on losing competitors, as described in the section on "Spoken-word identification." There, we address the question of exactly how lexical stress is processed in spoken-word recognition: does *camper* compete with *campaign* until the second syllables become segmentally unique, or does the stress difference in the initial syllable result in one alternative being ruled out before the incoming second syllable is heard?

Given the competitive nature of the lexical identification process, it would seem rational to assume that if the speech signal contains information that would contribute to the identification of the signal content, listeners would always make use of it. When acoustic cues to stress are available only as a function of higher-level structure, as in the fixed-stress case, the cues may be considered unreliable; but, for lexical stress, cues should be consistently available. Note, however, that in other speech-perception cases listeners do not always exploit all of the information available in speech. Even at the phoneme level, a contrast that is present (with the same acoustic effects) in many languages is not always resolved in the same way. A telling example concerns the two fricatives [f] and [s], contained in the phoneme repertoires of many languages; the mechanics of speech ensure that comparable local cues will be available whenever these sounds are uttered. However, listeners use such cues only in identifying [f] or [s] when the sound in question has a highly similar competitor in the language's repertoire (thus, only for identifying [f] in English and Castilian Spanish, because each of these languages has [θ] among its phonemes as well, and [θ] is highly confusable with [f]; but, in contrast, only for identifying [s] in Polish because Polish has [ɕ], [ʂ], and other sibilants: Wagner, 2013; Wagner, Ernestus, & Cutler, 2006). That is, cues may be consistently present, but language-specific phonological structure leads listeners to make use of them in some languages but to ignore them in others.

In the section on "Lexical stress and the vocabulary" we consider issues of vocabulary structure that could influence the perceptual relevance of lexical stress in a similar way. Fixed-stress languages clearly vary in the extent to which their phonology and vocabulary encourage any perceptual role for word stress (e.g. as described above for Polish, Hungarian, French and Finnish). Do lexical-stress languages vary likewise? Are there then also implications for the task in which lexical stress (but not fixed stress) can play a direct role, namely spoken-word recognition?

Lexical stress and the vocabulary

Every language has a store of sound-to-meaning pairings, that is, a vocabulary full of words, even though there are huge differences in the ways in which those stored elements map to parts of the speech signal (Fortescue, Mithun, & Evans, 2017). Among the dimensions along which vocabularies differ is the potential for interword competition in speech perception, and this is particularly determined by the size of the language's phonemic repertoire, and the constraints that the phonology applies to syllable structure. Each of these obviously affects the shape of the words that make up the vocabulary stock. Syllable structure includes whether and in which position consonant clusters can occur; a language in which syllables can vary from *oh* and *go* to *screech* and *plunged* will clearly allow more monosyllabic possibilities than a language in which only consonant–vowel (CV) and/or CVC syllables are legal. Segmental repertoire size is important in the same way, in that the more phonemes a language has, the more different short words its vocabulary can hold. The phonemic and syllabic properties of a language's vocabulary thus have direct implications for speech perception.

Consider first the word-length issue: shorter average word length in languages with more complex syllables and more phonemes, and longer average length in languages with less complex syllables and fewer distinct phonemes. This asymmetry certainly holds across the vocabularies of British English (45 phonemes) and Spanish (25 phonemes), for example; the mean word length in phonemes in English is 6.94, in Spanish 8.3, with length in syllables being respectively 2.72 and 3.48 (Cutler, Norris, & Sebastián-Gallés, 2004). Segmentally alone, there will be less embedding of short words within longer ones in the languages with larger phoneme repertoires and on average shorter words. Tallying word embedding and overlap in the vocabulary provides an estimate of the strength of competition from potentially coactivated candidate words in speech recognition, that is, spurious lexical competitors which can be activated by speech signals and impact upon spoken-word recognition. Indeed, the Spanish–English comparison by Cutler et al. (2004) showed this asymmetry too, with Spanish having more than twice the embedding rate of English.

The embedding rate can change if stress placement is considered as well. The potential for lexical stress to play a part in spoken-word identification can then be estimated by comparing the number of coactivated candidate words (effectively, the amount of competition) using the tallying just described while (1) taking only segmental structure into account, or (2) effectively including suprasegmental structure as well. Since these vocabulary statistics are computed using phonetic transcriptions, which represent segmental strings, adding this further dimension is realized by including the location of primary stress in each word's transcription, and ruling out any embeddings where syllables mismatch on this factor. On segmental match only, *enterprise* contains the shorter words *enter* and *prize*, and *settee* has *set* and *tea*. But if we require primary stress location to match also, *ENterprise* contains only *enter*, and *setTEE* contains only *tea* (neither *set-* nor *-prise* have primary stress in the longer words). Including the frequency of each carrier word as a weighting factor to estimate actual occurrence in speech experience, Cutler et al. (2004) found the difference between the two tallies to be much larger in Spanish (2.32 vs. 0.73) than in English (0.94 vs. 0.59). Thus, considering stress reduces competition in Spanish by more than two thirds, but in English by less than one third. Perhaps more importantly, on average a word will likely activate just a single competitor in English whether or not stress placement is taken into account, but in Spanish consideration of stress placement reduces more than two competitors to one competitor at most – a quantum improvement.

Interestingly, closely related languages such as English, Dutch, and German, all from the West Germanic family, can also differ on such a metric. The figures for these languages show that Dutch and German have more embeddings than English (Cutler & Pasveer, 2006). Examples of embedding in Dutch are *OUderdom* "old age," which contains *ouder* "parent" and *dom* "stupid," and in German *SAUna* "sauna" which contains *Sau* "sow" and *nah* "near." The effect of taking stress into account in identifying words, estimated in the same way as for English and Spanish, reveals that the amount of competition from embedding reduces by more than 50 percent for both Dutch and German if stress match is included in these computations. With carrier-word frequency taken into consideration, a Dutch segments-only count then gives 1.52 competitors per word of speech on average, and a segments-plus-stress count reduces this to 0.74; for German, the respective figures are 1.72 and 0.8. This is again a quantum improvement (from more than one to less than one competitor) for each language, which English cannot match because its embedding count was below 1 even without taking stress into account. The differences have implications for word recognition in these languages.

For instance, the fact that taking stress marking into account produced no substantial improvement in English could imply that English listeners actually need to attend only to words' segments in computing mismatch to competitors; vowels and consonants reduce competition sufficiently for optimal or near-optimal recognition. Adding the use of suprasegmental cues would thus not reduce competition to an extent that would be worth the effort. This suggestion assigns a vital role to the segmental correlate of stress, the central vowel schwa. In languages where schwa cannot be stressed, the presence of schwa effectively signals lack of stress without the need of another cue dimension. The relative rarity of embeddings in English compared with Dutch and German would reflect the greater likelihood of an English unstressed syllable containing schwa rather than a full vowel. Can the vocabulary provide specific evidence on this proposal?

A comparison of English and Dutch shows that both languages have extensive affixation, with the affixes (especially inflectional suffixes, but also prefixes) being typically realized by weak syllables containing schwa in each language. Over the entire vocabulary, morphological differences between the two languages actually tend to produce more Dutch than English syllables with schwa (in both initial and final position). Neither in initial nor in final syllables do the vocabulary counts show greater schwa frequencies for English. Strong effects of schwa are found instead in nonaffixed syllables, which can occur in word-medial positions. Comparisons within morphological families show less phonetic overlap in English – in *admire admiration*, *gratitude gratuity*, *legal legality*, and so on, stress alternation brings different vowel realization with it, so that the initial syllables do not compete; specifically, the position of schwa switches between the two pair members. There is more phonetic overlap in Dutch: pairs such as *legaal legaliteit* "legal legality," *glorie glorieus* "glory glorious," and *definitie definitief* "definition definitive" all share the segments of the initial bisyllables despite having differently placed primary stress. Thus syllables without primary stress more often contain schwa in English than is the case for Dutch, and unstressed syllables with full vowels are more common in Dutch than in English.

The vocabularies reveal different, indeed opposing, patterns; taking into account all words of three or more syllables, Dutch has more full vowels in medial syllables without stress, while English has more schwa. Despite the lack of overlap in morphologically related pairs, English has more overlap in general. A tally of the proportion of words in the vocabulary that share an initial bisyllabic string with one or more other words reveals a significantly larger figure for English than for Dutch (Bruggeman & Cutler, 2016; consider that, for example, the initial CVCV sequence of the English words *coral*, *correlate*, *corridor*, *coroner*, *corrugated* and *coryphée*, although spelled differently in each word, is in each case phonetically identical, the constant second V being schwa). The effect of such greater overlap is to increase competition, of a kind that taking stress into account is not going to help at all.

Finally, the position of embedded words within their carriers also varies across vocabularies. In the Germanic languages, embedded words in carrier-initial position way outnumber embeddings in final position. A combination of suffixal morphology and vowel reduction in unstressed syllables causes this skew, which therefore hardly exists in some other languages. Consider Japanese, which, in contrast to English, Dutch, and German, is a noninflectional language, and has neither stress nor vowel reduction; it has no significant asymmetry of embedding position. Because both Dutch and German have more inflectional suffixes (on verbs and as noun plurals) than English, the tendency toward more initial than final embeddings, which is already larger in English than in Japanese, is significantly larger again in Dutch and German. The added fact that German has very many monomorphemic words ending in schwa finally boosts German to the largest asymmetry in the set (Cutler, Otake, & Bruggeman, 2012).

These comparisons might suggest that both morphology and stress phonology are necessary to explain the initial-final embedding asymmetry, but in principle either could actually produce it alone. Evidence from the vocabulary of French offered a further explanation here; French has suffixes aplenty in its morphology and schwa in its phoneme repertoire, but no lexical stress in its phonology. Comparison of French with the other vocabularies (Cutler & Bruggeman, 2013) revealed that the suffix effect alone, present in French, accounted for a positional asymmetry roughly half the size of that in English. Then, if the lexicon of French were modified to allow realization (as schwa) of the "silent" vowel in words such as *petite* "small" and *ville* "town," the asymmetry roughly equaled that of English. Thus morphology on the one hand, and segmentally realized stress on the other, each separately and additively contribute to this influence on the availability of competing words during speech processing.

It is clear, therefore, that vocabulary analyses such as those reviewed in this section predict differences across lexical-stress languages (even among those most closely related) in the degree to which the processing of suprasegmental information leads to a worthwhile payoff in the efficiency of spoken-word recognition. The next section reviews the findings relevant to this proposal.

Spoken-word identification

The vocabulary statistics are based on measures of overlap, on the assumption that interword competition is the primary testbed for whether a particular factor will play a useful role in the identification of spoken words. Given that these measures differ across lexical-stress languages when suprasegmental cues are considered, the statistics then predict cross-language differences in whether suprasegmental cues will actually prove useful to listeners.

Evidence for the importance of the competition factor on which the statistics are based is firmly established. Competitor words, temporarily supported by an incoming speech signal, are effortlessly discarded by listeners as mismatching speech information becomes available, but traces of their fleeting presence are reliably seen in psycholinguistic experiments. For instance, in the cross-modal priming task, where listeners make yes–no lexical decisions about printed words while hearing speech, words are recognized more slowly when they partially match the auditory input than when the auditory input is totally unrelated (e.g. responses to printed *feel* are slower after spoken *feed* than after spoken *name*; Marslen-Wilson, 1990), and in the word-spotting task, finding real words in nonsense carriers is harder if the carrier could become another real word than if it could not (so *WRECK* is easier to find in *berrec* than in *correc*, which could be the beginning of *correction*; McQueen, Norris, & Cutler, 1994). This response inhibition in each case is evidence of the temporary availability but later rejection of a different interpretation of the speech input. Candidate words that have been subject to competition are momentarily less available to the recognition process. Initial competition has greater effects than final competition on the speed and accuracy of word recognition too; this asymmetry between embeddings in word-initial versus word-final position has been supported in many spoken-word recognition studies in English

(Allopenna, Magnuson, & Tanenhaus, 1998; Cluff & Luce, 1990; McQueen & Viebahn, 2007).

Exactly as the statistics predict, the relative usefulness or otherwise of suprasegmental cues to stress differs across languages. In Dutch, listeners' guesses, given increasingly larger fragments of pairs such as *CAvia kaviAAR* "guinea pig caviar" in a gating task, were clearly shown to draw on suprasegmental as well as on segmental information (van Heuven, 1988). Mis-stressing of Dutch words likewise exercised adverse effects on recognition both in gating (van Leyden & van Heuven, 1996) and in semantic judgment tasks (Koster & Cutler, 1997). Single syllables differing only in stress could be assigned to the appropriate source word by Dutch listeners (e.g. the above *CA-/ka-* pair; Cutler & van Donselaar, 2001; Jongenburger, 1996) and also by German listeners (e.g. *AR-/ar-* from *ARche ArCHIV* "ark archive"; Yu, Mailhammer & Cutler, 2020). Spoken-word recognition studies in English, in contrast, have repeatedly failed to find equivalent effects. Studies in English showed that mis-stressing of stress pairs did not at all affect their recognition (Small, Simon, & Goldberg, 1988), that mis-stressing also did not affect word recognition in noise (Slowiaczek, 1990); that minimal stress pairs such as *TRUSty* and *trusTEE* activate both associated meanings just as homophones would (Cutler, 1986), that stress pattern prompts were ignored in word-matching judgments (Slowiaczek, 1991), and that cross-spliced words in which primary- and secondary-stressed syllables were swapped were rated as just as natural as the original unmodified versions (Fear, Cutler, & Butterfield, 1995). All these findings could be said to give evidence of the low payoff provided by the English vocabulary for the use of suprasegmental as well as segmental information in identifying English words, while the Dutch results confirmed the greater utility of paying attention to both cue types in that language.

Note that these results largely made use of "offline" tasks, that is, they did not tap word-processing speed, but only whether or not the outcome was correct. Cross-modal priming studies that measure word-recognition speed, however, likewise revealed language-specific differences in the use of cues to stress in the process of recognizing spoken words. As described earlier, the cross-modal priming task critically allows a view of competition, via an inhibitory effect on the recognition of a constant target when competition is present versus when it is absent. Evidence from a series of studies using cross-modal fragment priming (where the primes were fragments of spoken words) indeed indicated that suprasegmental information about lexical stress could modulate lexical processing. In these studies, respectively in Spanish (Soto-Faraco, Sebastián-Gallés, & Cutler, 2001), Dutch (van Donselaar, Koster, & Cutler, 2005), and English (Cooper, Cutler, & Wales, 2002), listeners heard a fragment prime taken from the onset of longer word pairs in their respective native language, before performing a lexical decision on a printed target. A study in German (Friedrich et al., 2004) combined this cross-modal priming technique with a record of evoked response potentials (ERPs) in the brain.

In all these studies, the fragment in matching conditions consisted of the initial portion of the target word. That is, two-syllable fragments had the same segmental composition and stress pattern as the first two syllables of the target word (e.g. in the case of English, *ADmi-* was a prime for *ADmiral*). Listeners' lexical decisions were faster in this matching condition, compared in the Dutch, Spanish and English studies to a control condition with unrelated primes such as *immer-* from *immersion,* and in the German study, which had no control-prime condition, to the mismatching-prime condition. Facilitatory priming was also found when primes were monosyllabic (Cooper, Cutler, & Wales, 2002; Friedrich et al., 2004; van Donselaar, Koster, & Cutler, 2005). In the German ERP study, mismatching fragments of any size further induced a positive response approximately 350 milliseconds following the onset of target presentation (a P350) which was held to be a signal of the detection of lexical incongruency.

Such facilitation over a control is expected on the basis of segmental mismatch alone, of course; the critical information about whether suprasegmental cues to lexical stress contribute to word recognition is provided by comparing the control to a stress-mismatching condition. In this condition, the fragment primes also matched the first portion of the target word in their segmental composition, but mismatched them in stress (e.g. *ADmi-* before *admiRAtion*). Stress in these words was solely cued through suprasegmental information, as the vowels in matching and mismatching primes were identical. Here the results depended on the language: Dutch and Spanish listeners responded more slowly to target words after a fragment prime with a different stress pattern, compared to after an unrelated prime. For these language users, the suprasegmental differences in stress pattern weighed enough to undo the advantage that the match in segmental overlap had provided. When fragment primes were monosyllabic (a condition in the study with Dutch listeners, but not in the Spanish study) this inhibitory effect was not observed. The monosyllabic primes thus gave Dutch listeners insufficient suprasegmental information to favor stress competitors over the target words. Together, the results show that Dutch and Spanish listeners can use suprasegmental information about lexical stress in the recognition of words, although they do so more effectively if two syllables are provided. In Dutch and Spanish, both segmental and suprasegmental information thus provide support for word representations during recognition.

In contrast, the results again showed English listeners to be less efficient in using the suprasegmental cues to lexical stress in lexical processing (Cooper, Cutler, & Wales, 2002). While English listeners' responses had also been facilitated in the matching conditions, their results for the stress-mismatching conditions differed from those observed for Dutch and Spanish listeners. Bisyllabic stress-mismatching primes crucially did not inhibit recognition. They also did not elicit facilitatory priming, however, suggesting that the relative contributions of the segmental match and the suprasegmental mismatch may have offset one another. English listeners, like Dutch listeners, also obtained less stress information from monosyllabic primes than from bisyllabic primes; indeed, they showed facilitation for stress-mismatching monosyllabic primes. In English word identification decisions, suprasegmental information about lexical stress may therefore be outweighed by segmental information. Alternatively, as a result of listeners' experience of the English vocabulary structure reviewed earlier, suprasegmental cues may simply be less efficiently processed than segmental cues.

Note, though, that the stress distinction in the English study may also be held to have been intrinsically less useful to listeners. In the English study, bisyllabic primes differed in whether they had primary or secondary stress on the first syllable, with the second syllable being constantly weak (with schwa). Such contrasts can be obtained in English, but there are no contrasts such as those between the full vowels in the first two syllables of Dutch *OCtopus* and *okTOber* or Spanish *PRINcipe* "prince" and *prinCIpio* "beginning," in which the vowels remain the same although the primary stress shifts from first to second syllable. In English *OCtopus* versus *ocTOber,* the vowels are not the same (the medial syllable of *OCtopus* has schwa) so that for this reason sufficiently there is no competition. Half the word pairs in the Dutch study were similar in structure to the English pairs, but the remaining half offered listeners a less subtle and hence potentially more useful contrast. Again, what the vocabulary offers in Dutch (or Spanish), compared to English, is intrinsically more useful to listeners than what the English vocabulary can provide.

In summary, these priming studies confirm the "offline" findings that suprasegmental cues to lexical stress can aid spoken-word recognition, as long as the statistical payoff provided by the vocabulary is significantly strong. However, cross-modal priming studies also suffer from a limitation: they cannot inform researchers whether such cues are evaluated rapidly and efficiently enough to facilitate spoken-word recognition in continuous speech. In cross-modal priming, the response measure is time to accept that the printed target is indeed a real word, with the visual presentation being a separate operation from the auditory processing. Another experimental paradigm that combines visual and auditory processing, but in a dependent rather than disjointed manner, is eye tracking, in which participants are instructed to look at a visual display and their looks are tracked. As speech unfolds over time, information about a spoken word gradually becomes available and constrains its identity, and if the display contains a picture (Allopenna, Magnuson, & Tanenhaus, 1998) or a printed string (McQueen & Viebahn, 2007) corresponding to an incoming spoken form, looks to that part of the display will accumulate as a function of the information availability. In other words, this technique offers a way of tracking the uptake of speech input across time. Thus the priming and eye-tracking methods offer complementary views of the spoken-word recognition process: eye tracking reveals what speech information is available when, while priming provides a record of what active (inhibitory) competition has occurred. The full picture requires both sources of information.

Looking behavior does allow tracking of the degree to which other candidate words are considered over time, of course; listeners immediately respond to even fine-grained sub-phonemic information by adjusting their looking patterns (Dahan et al., 2001; McMurray, Tanenhaus, & Aslin, 2002). The proportion of looks to competitors in a display shows the degree to which listeners simultaneously consider multiple candidates (and potential candidates are not necessarily limited to words in the current display; Dahan et al., 2001; Magnuson et al., 2007). Such effects on looking patterns likely result from active lexical competition, even if the task offers no direct measure of the consequent inhibition.

Recent eye-tracking work has shown listener entrainment to prosodic structure (Brown et al., 2015), and studies in three languages have confirmed that suprasegmental cues to lexical stress can influence spoken-word recognition (Jesse, Poellmann, & Kong, 2017; Reinisch, Jesse, & McQueen, 2010; Sulpizio & McQueen, 2012). Participants in these three experiments heard sentences containing a member of a critical stress pair (e.g. *Click on the word ADmiral*) while the fixations of their eyes on a computer screen were tracked. The display would then include the printed target word (*admiral*), its stress competitor (*admiration*), and two unrelated words; participants simply followed the instructions to click on the mentioned word. The first two syllables of the competitor overlapped segmentally with the target word, but differed suprasegmentally, with words' target or competitor roles counterbalanced across trials. The eye-movement data showed that Dutch, Italian, and English listeners fixated on target words (in their respective native language) more frequently than competitors even before disambiguating segmental information was available.

Listeners from all these three languages can thus process suprasegmental cues to lexical stress efficiently enough to facilitate recognition of the spoken word. Relative efficiency of processing segmental compared to suprasegmental cues as an explanation of the English processing results can therefore be rejected. Nonetheless, cross-language differences in the contribution of suprasegmental information were once again observed. In Dutch (Reinisch, Jesse, & McQueen, 2010), there was an effect of stress pattern in that competitors with primary stress on the initial syllable attracted more looks than competitors with secondary stress, and in Italian (Sulpizio & McQueen, 2012) there was an effect of default stress pattern (to be described in detail); but in the English experiment (Jesse, Poellmann, & Kong, 2017) and also in a replication by Kong and Jesse (2017) there was no sign of a stress pattern effect. As argued by Jesse and colleagues, this pattern suggests that, while cue processing can be equally efficient across languages, English listeners attach greater weight to the segmental than to the suprasegmental information they receive, an imbalance that is absent from processing by the Dutch or Italian listeners. We return to this proposal in the conclusion to the chapter.

It therefore appears that, with sufficiently sensitive techniques, evidence of efficient suprasegmental stress cue processing can always be observed. The striking differences across languages in many studies and their obvious links to vocabulary structure signal that stress perception itself is a different issue from whether stress is useful, and for what.

Note that there can be many ways in which native listeners show that they have efficiently registered their language's stress features. In the case of English, the most telling evidence comes from the speech segmentation literature, where the predominance of stress-initial words in English speech (Cutler & Carter, 1987) is the presumed underpinning of the fact that listeners who are locating word boundaries in that language use stressed syllables as indicating the beginning of a new word (Cutler & Norris, 1988). Also, word-class stress regularities (final stress in English bisyllables is significantly more common for verbs than for nouns) form part of stored lexical representations (Arciuli & Slowiaczek, 2007). Indeed, English speakers exhibit knowledge not only of the major stress placement probabilities, but also of subtle stress-led probabilities such as that trisyllabic words with [i] in their final syllable are more likely to bear antepenultimate (*cavity*, *recipe*) than penultimate stress (*safari*, *bikini*; Moore-Cantwell & Sanders, 2017). Further, they are capable of learning novel rules based on stress, such as consonantal occurrence restrictions that are dependent on trochaic or iambic stress contexts (White et al., 2018). Stress errors in speech production are not overlooked by listeners, but induce misperceptions (Bond & Garnes, 1980; Cutler, 1980; Fromkin, 1976). All these abilities rest on perceptual processing of stress, which is not signaled by segmental structure alone.

Indeed, participants in English word recognition experiments can be seen to be making use of suprasegmental cues to stress; they just do not always do so as efficiently as listeners from other lexical-stress language communities. Cooper, Cutler, and Wales (2002), besides their cross-modal priming studies, conducted a simpler "offline" task in which listeners heard word-initial syllables extracted from pairs of words with segmentally identical but suprasegmentally differing initial syllables (such as *hum-* from *HUmid* versus *huMANE*), and chose which member of the pair had in each instance been the source word. Mattys (2000) had also performed such a task with initial syllables of trisyllabic pairs such as *PROsecutor* versus *proseCUtion*. In neither study were such pairs of single syllables reliably distinguished at an above-chance level, and in Cooper et al.'s data, the most striking finding was that Dutch listeners proficient in English outperformed the native English listeners with the English materials, in particular by correctly identifying noninitially stressed cases such as *hum-* from *huMANE* (where the English native listeners' responses did not differ from chance). Later work showed that German listeners proficient in English likewise performed above chance in correctly classifying these noninitially stressed English word fragments (Yu, 2020). Further, a similar judgment task in Spanish revealed that English listeners failed to use the F0 and durational cues to Spanish stress placement to the extent that native Spanish listeners did (Ortega-Llebaria, Gu, & Fan, 2013).

However, more detailed analyses (Cutler et al., 2007) of Cooper, Cutler, and Wales's (2002) English data revealed that those listeners had indeed made use (if somewhat inefficiently) of stress cues. Measurements showed that, although the pairs differed significantly in the initial syllables' duration, F0, amplitude, and spectral tilt, it was the F0 cues that most reliably distinguished the pairs (separate evidence from a gating task with Dutch listeners had also proved the F0 cues to be most informative). Correlations across item means then showed that Cooper et al.'s Dutch listeners made good use of each type of cue, in that the degree of acoustic difference between pairs correlated with the likelihood of a correct decision. However, the results for the native English-speaking listeners differed: the F0 cues were appropriately exploited, but the weaker cues were not. In fact, some correlations seemed hard to explain in that they were in the opposite direction from what might have been predicted; unstressed syllables were significantly shorter, and had less spectral tilt, than stressed syllables, but the longer such a syllable was and the greater its spectral tilt, the more likely English listeners were to correctly judge it as unstressed. Considering that longer duration, and clearer information in the upper spectral ranges, should each have increased listeners' opportunity to process F0, it may be that these paradoxical results are actually cues to what the English listeners were doing: correctly exploiting (only) the most effective of the stress cues, that is, F0. When placed in a situation where the only information relevant to performing a task is information that they don't often use in this way, English listeners can do their best and at least exploit the clearest and most reliable cue.

In other languages, too, stress perception abilities can be drawn upon in the appropriate circumstances, but otherwise not. The Italian eye-tracking study of Sulpizio and McQueen (2012), confirming cross-modal priming findings by Tagliapietra and Tabossi (2005), showed that Italian listeners could certainly use suprasegmental stress cues in recognizing words. However, they did not use them for every word; they used them only for those experimental stimuli that had antepenultimate stress (e.g. *COmico* "funny"). This may seem arbitrary, but in fact is not, since this stress pattern is an exception to the general stress rules for Italian. With words in which the more common default rules applied (e.g. *coMIzio* "meeting"), listeners ignored the suprasegmental cues. Relatedly, a study in Turkish (Domahs et al., 2013) measured ERPs to examine listeners' responses to hearing correctly stressed words versus incorrect stress realized by manipulating F0, duration, and amplitude together; note that ERP evidence confirms that Turkish stress minimal pairs such as *BEbek* "a district of Istanbul" vs *beBEK* "baby" can be distinguished on these suprasegmental cues alone (Zora, Heldner, & Schwarz, 2016). In Domahs et al.'s study, different types of violations led to different responses. The expected default stress placement in Turkish is final, and violations of default stress modulated a P300 effect. In contrast, violations of stress on words with an exceptional stress placement (e.g. initial), that had to be stored as part of their individual phonological representations, produced an N400 effect. Default stress patterns are therefore processed differently than lexically defined exceptions, in Turkish (classified as fixed stress given the predominance of the default case) as in Italian (classified as lexical stress since predominant patterns differ with word length).

This ERP research has thus established that listeners from a fixed-stress language background can also process suprasegmental cues when necessary. Further, though users of fixed-stress languages show stress "deafness" by being unable to recall stress placement long enough to perform an ABX choice, they can correctly perform the simpler AX discrimination (same–different, e.g. *bopeLO–boPElo;* Dupoux et al., 1997). Vroomen, Tuomainen, and de Gelder (1998) likewise showed that Finnish listeners could use F0 cues to word-initial stress in learning the "words" of an artificial language; and word-final suprasegmental cues are exploited in continuous-speech segmentation by listeners from the fixed-final languages French (Tyler & Cutler, 2009) and Turkish (Kabak, Maniwa, & Kazanina, 1999). In Hungarian, an ERP study of suprasegmental cue processing by Garami et al. (2017) showed that oddball detection, as indicated by the mismatch negativity component, differed between real and pseudo-words; cues were processed more efficiently in pseudo-word than in real-word input. That is, when listeners knew that an incoming item was a real word, they evaluated the stress cues differently; indirectly, this shows that they could hear and process cues to stress.

The way that listeners process these cues is thus in no way dependent on language-specific perceptual ability; rather it is language-specific vocabulary training that determines the attention that stress cues receive. This is as true for fixed- as for lexical-stress languages.

New horizons for stress in speech perception

As the previous section revealed, new techniques have caused recent research in speech perception to substantially deepen our understanding of the role of stress in the processing of spoken words. We expect that this trend will continue. ERP research in particular will expand in this field, with lexical processing of suprasegmental information already having been shown to distinguish between minimal-pair competitors (in Swedish, where accents can signal word number: Söderström et al., 2016). New fields have also opened perspectives from which word stress can be further understood. This includes applied fields such as the processing of prosody in second-language speech perception and production, which was long a virtually unresearched topic but is now the subject of a rapidly expanding literature that has already grown too extensive to include here. In this section, we report briefly on two growing areas that shed additional light on the central question of the preceding section: How does word stress contribute to word recognition? These concern speech perception when listening has become difficult because the input is degraded, and speech perception considered as an audiovisual process.

Lexical-stress perception in degraded speech

Suprasegmental cues to lexical stress can facilitate listeners' recognition of spoken words, by providing more support for the target word over its lexical competitors. This facilitation depends, however, on the quality of the speech signal. Mattys, White, and Melhorn (2005) have argued that word stress may become most relevant in noisy listening situations, such that the use of word stress for the segmentation of continuous speech increases when lexical and semantic cues are degraded by noise. However, when the quality of the available stress cues themselves is artificially degraded, listeners may default to the pattern most common in their native language. In a second experiment in their study described above, Sulpizio and McQueen (2012) taught Italian listeners to associate nonsense shapes with novel nonwords, with stress on the penultimate (e.g. *toLAco*, the default pattern) or on the antepenultimate syllable (*TOlaco*, the exceptional pattern). Amplitude and duration cues were neutralized in the training nonwords, while F0 cues remained intact. In a test phase, participants were then instructed to click on a particular shape, and their fixations on that shape or shapes associated with a stress competitor or a distractor were tracked. During this test, the nonwords had only F0 stress cues, as in training, or were fully intact. With only F0 cues, nonwords with penultimate stress competed more for recognition than those with antepenultimate stress. That is, with insufficient acoustic support for a definitive answer, participants showed a general preference to interpret the nonwords as bearing the default stress. Knowledge of vocabulary patterns is presumably helpful in difficult listening conditions outside the laboratory as well.

One case in which suprasegmental cues to prosody are definitely degraded is when listening occurs through cochlear implants (CIs). In particular, the spectral and temporal fine-structure information needed for the perception of pitch is reduced in cochlear-implant listening, for example, as it is discarded by the signal-processing algorithms implemented in CIs (Smith, Delgutte, & Oxenham, 2002; Zeng et al., 2005). CI users thus perceive prosodic contrasts less well than normal-hearing listeners (e.g. Holt & McDermott, 2013; Holt, Demuth, & Yuen, 2016; Meister et al., 2009; Morris et al., 2013; Peng, Lu, & Chatterjee, 2009). Swedish normal-hearing listeners are therefore better, for example, than CI users at recognizing words from minimal stress pairs in a two-alternative forced choice (Morris et al., 2013). This performance difference was exacerbated when speech-shaped noise was added, as then only the performance of the CI users declined.

Among CI users, those with residual low-frequency hearing often perform better in perceptual tasks than those without residual hearing (Ching, van Wanrooy, & Dillon, 2007; Gifford et al., 2012; Woodson et al., 2010), including on tasks dependent on the processing of suprasegmental information (Kong, Stickney, & Zeng, 2005; Zhang, Dorman, & Spahr, 2010). For example, both CI users with and without residual hearing use stressed syllables in English as cues to word onsets in segmenting continuous speech (Spitzer et al., 2009). However, only the performance of CI users with residual hearing changed as a function of whether or not pitch was provided as a cue. Low-frequency residual hearing improves the perception of pitch (e.g. Dorman et al., 2008; Kong, Stickney, & Zeng, 2005), and as such may enhance CI listeners' ability to facilitate speech perception through prosodic information.

While these results suggest that residual hearing provides a benefit for discriminating prosodic contrasts, or even enables it, these studies cannot speak to how prosodic cues in speech input are processed by CI users, in particular whether or not any CI users can process prosodic cues effectively enough for these to have an immediate influence on ongoing lexical processing. To address this question in the case of lexical stress, Kong and Jesse (2017) simulated the degradation of speech in CI listening by using (eight-channel) noise-vocoded speech. Normal-hearing participants were first trained to recognize spectrally vocoded speech, before being tested, in the eye-tracking paradigm also used by Jesse, Poellmann, and Kong (2017), on vocoded speech with and without supplementary low-pass filtered speech information. Critical word pairs (e.g. *Admiral admiRAtion*) differed again in primary or secondary stress on the first syllable, with unstressed, unreduced second syllables being identical across a pair and segmental difference not appearing until at least the third syllable.

In the vocoder-only condition, participants could not distinguish target from competitor words using lexical stress information alone. In Jesse, Poellmann, and Kong's (2017) study with intact speech, only pitch and amplitude differed across the word-stress patterns; but here noise-channel vocoding discarded the fine-structure information needed to access pitch, and the remaining amplitude cues did not suffice. There was also no sign that listeners had applied a general strategy, for example choosing word-initial stress, as the most common pattern for English words. In contrast, in the condition where vocoded speech had been supplemented with a low-pass filtered version of the speech materials, listeners showed that they were able to effectively use suprasegmental cues to lexical stress to determine the target word. Importantly, these differences across listening conditions were not based on differences in access to segmental information. The preference for fixating a competitor word over phonologically unrelated distractor words was the same across listening conditions, indicating a similar degree of lexical competition due to segmental overlap.

Lexical stress in visual speech

In many of our daily conversations, we hear and see the speaker. It is well established that listeners use available information from both modalities to achieve more robust recognition of speech (e.g. Jesse & Massaro, 2010; Jesse et al., 2000; MacLeod & Summerfield, 1987; Sumby & Pollack, 1954). While most work in the domain of audiovisual speech perception has focused on how visual speech aids recognition by providing segmental cues, visual speech can also provide prosodic information (e.g. Bernstein, Eberhardt, & Demorest, 1989; Dohen et al., 2004; Munhall et al., 2004; Srinivasan & Massaro, 2003). Minimal word-stress pairs can, for example, be distinguished based on speech-reading alone. Thus Risberg and Lubker (1978) showed that Swedish normal-hearing and hearing-impaired adolescents can (equally well) distinguish minimal pairs that differ in stress placement (e.g. *BAnan* "rack" vs. *baNAN* "banana"). Likewise, English adults in a speech-reading study by Scarborough et al. (2009) performed above chance when asked to distinguish noun–verb stress pairs (e.g. [a] *SUBject* vs. [to] *subJECT*) as well as reiterant speech versions of these minimal pairs (e.g. *FERfer* vs. *ferFER*). Together, these results suggest that lexical stress has visual correlates that listeners can use in spoken-word recognition. For instance, in Scarborough and colleagues' study, stressed syllables were produced with a larger lip opening and with larger and faster chin movements than unstressed syllables with reduced vowels. However, it is unclear which of these visual correlates perceivers relied on.

Degrees of lexical stress can also be recognized from visual speech. Presenting the first two syllables of Dutch word pairs as visual speech, Jesse and McQueen (2013) showed that participants could distinguish primary from secondary stress preceding an unstressed syllable (e.g. *CAvi-* from *CAvia* "guinea pig" vs. *kavi-* from *kaviAAR* "caviar") and unstressed-primary from secondary-unstressed stress sequences (e.g. *proJEC-* from *proJECtor* "projector" vs. *projec-* from *projecTIEL* "projectile"). The former distinction was, however, possible only when the critical fragments came from words with phrase-level emphasis. Phrase-level emphasis falls onto syllables with primary stress and increases articulatory effort (de Jong, 1995; Fowler, 1995; Kelso et al., 1985). Cues in primary-stressed syllables were then strengthened by the phrase-level emphasis. The otherwise subtle difference between the visual correlates of suprasegmental differences across syllables thus became perceptible.

Together, these studies demonstrate that information about the lexical stress pattern of a word can also be obtained from seeing a speaker. Perceiving lexical stress from visual speech is not reliant on segmental cues, but rather on the visual correlates of suprasegmental cues to lexical stress, and these can be sufficient for recognition – at least in languages where word recognition relies more heavily on such cues.

Conclusions

Word stress is not language universal, and even among languages that do mark stress within words there is variety: the placement of stress can be fixed, or it can be variable. If it is fixed, the stipulated position may be at a word's edge or not, with different implications for a demarcative function for stress; if it is variable, it may still be affected by phonological factors such as the availability of vowel reduction. Either type of stress may show traits of the other too: fixed stress languages may have some minimal pairs involving loan words or proper names; lexical stress languages can display strong tendencies toward preferred positions for the placement of primary stress. All these features are captured in each language's vocabulary, and, as we have argued, it is via learning of a vocabulary that listeners come to know how they should make use of cues to stress in speech perception.

The study of stress in speech perception is expanding, and we particularly look forward to data from languages previously unrepresented in this literature. For now, however, our review has largely drawn from English and related languages. Even here we find plenty of data to establish our argument that the use of speech cues to stress is driven by vocabulary structure. We drew particular attention to subtle mismatches in the processing of suprasegmental cues to stress in the closely related languages Dutch and English. If stress location is ignored, the Dutch vocabulary shows a higher degree of within-word embedding than the English vocabulary, but if the computation takes stress location into account this asymmetry is greatly reduced, due to significant reduction in the Dutch embedding counts. The vocabulary pattern thus suggests that Dutch listeners will be able to speed spoken-language processing by attending to where stress falls to a greater degree than English listeners. This prediction offers an explanation for a remarkable asymmetry in the processing data; studies of the uptake of suprasegmental cues to stress show that both English and Dutch listeners can process these, and equally efficiently, but studies focusing on the resolution of active competition, by inhibition of competing forms, reveal that such inhibition via suprasegmental cues occurs only in Dutch. English and Dutch listeners can both use suprasegmental cues to help identify words, but English listeners do not further use the same cues to suppress competitors, while Dutch listeners do.

We have argued that this asymmetric pattern appears because, in English, with its lesser degree of embedding, stress information has a lower payoff in the word-identification process. Listeners can afford to let suprasegmental information regarding word identity be outweighed by segmental information. Note that the relative weighting of different sources of information in lexical access is a well-studied topic; many lines of research have demonstrated, for instance, that consonantal information yields stronger cues to lexical identity than vowel information does (see Nazzi & Cutler, 2019, for a review). This weighting effect for segments appears across languages differing widely in phoneme repertoire makeup, and is seen in findings such as listeners' greater willingness to turn nonwords (*shevel*, *eltimate*) into real words by changing a vowel (*shovel*, *ultimate*) than by changing a consonant (*level*, *estimate*; van Ooijen, 1996) or in the greater discoverability of "words" in an artificial language with consonant regularities than in one with vowel regularities (Mehler et al., 2006). There are orthographies (such as Hebrew) in which consonants may be written and vowels omitted, but there are no cases of the reverse pattern. The vowel/consonant asymmetry is also driven by the vocabulary; for instance, while young infants at first favor listening to vocalic over consonantal information (held to be due to the talker identity cues carried on vowels), this infant pattern of vowel preference switches to the adult pattern of consonant preference once the compilation of an initial vocabulary begins, toward the end of the first year of life (Nazzi & Cutler, 2019).

Vowel/consonant preferences relate directly to the perception of stress because suprasegmental cues to stress patterns are primarily carried on vowels: F0 movements are realized across vowels, amplitude is principally realized in vowel articulation, and vowels make a significantly greater contribution to syllable duration than do consonants. Thus downgrading the contribution of vocalic information, which has been established to happen in many languages, will automatically result in a corresponding downgrading of suprasegmental information. From this point of view, it may seem that these lexical weighting patterns, in which suprasegmental information contributes little to word identification, in fact constitute the default case. Indeed, among all the world's languages, those with lexical stress form a minority, outnumbered by the combination of fixed-stress languages and languages without word-level stress. Even among lexical-stress languages, however, suprasegmental cues seem to be fully exploited only when their use noticeably speeds processing by significantly reducing the lexical competitor count. The available option of modulating lexical competition by weighting all cues (segmental or suprasegmental) equally is chosen when (and only when) the vocabulary renders it profitable.

As the section on "Lexical stress and the vocabulary" showed, the size (and relative benefit) of this competitor reduction can be estimated by computing word overlap in the vocabulary in the light of frequency statistics for word usage. Reliable large corpora, especially of spoken language, so far exist for only relatively few languages; but if there is one thing that is growing fast worldwide, it is the collection and exploitation of large data sets, so the prediction is that in the future this particular data problem should be solved. Then vocabulary-based predictions about the use of suprasegmentals in lexical processing can be made, and subsequently tested, for many more lexical-stress languages.

We note that at the level that creates the general processing strategies at issue here, the phonological patterns characterizing a vocabulary are largely constant across varieties. For English, our reports suggesting that suprasegmental cues are downgraded in speech perception have come from different varieties (American English, e.g. Jesse, Poellmann, & Kong, 2017; British English, e.g. Fear, Cutler, & Butterfield, 1995; Australian English, e.g. Cooper, Cutler, & Wales, 2002), but they all paint the same picture. Indeed, crucial findings directly replicate across varieties (for instance, McQueen, Norris, & Cutler's [1994] demonstration of active interword competition was conducted in British English, but an identical pattern, including strong effects of word-stress placement, appears in American English: Warner et al., 2018). These varieties of English essentially share all their major vocabulary patterns, notwithstanding the existence of individual variety-specific lexical items. It is not impossible that varieties of one language will differ in the degree to which suprasegmental information is used in speech perception; but analysis of variety-specific dictionaries can quickly establish whether relative competition patterns predict this to happen.

Our knowledge of the processing of suprasegmental information in speech perception has significantly expanded in recent years. The dependence of processing patterns on the vocabulary is clear, and with it the path to yet further discoveries.

Acknowledgments

AC acknowledges support from the Australian Research Council Centre of Excellence for the Dynamics of Language (CE140100041).

REFERENCES

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language*, *38(5)*, 419–439.

Arciuli, J., & Slowiaczek, L. M. (2007). The where and when of linguistic word-level prosody. *Neuropsychologia*, *45(11)*, 2638–2642.

Berinstein, A. E. (1979). A cross-linguistic study on the perception and production of stress. *UCLA Working Papers in Phonetics*, *47*.

Bernstein, L. E., Eberhardt, S. P., & Demorest, M. E. (1989). Single-channel vibrotactile supplements to visual perception of intonation and stress. *Journal of the Acoustical Society of America*, *85*, 397–405.

Bond, Z. S., & Garnes, S. (1980). Misperceptions of fluent speech. In R. Cole (Ed.), *Perception and production of fluent speech* (pp. 115–132). Hillsdale, NJ: Lawrence Erlbaum.

Brown, M., Salverda, A. P., Dilley, L. C., & Tanenhaus, M. K. (2015). Metrical expectations from preceding prosody influence perception of lexical stress. *Journal of Experimental Psychology: Human Perception and Performance*, *41*, 306–323.

Bruggeman, L., & Cutler, A. (2016). Lexical manipulation as a discovery tool for psycholinguistic research. In C. Carignan & M. D. Tyler (Eds.), *Proceedings of the Sixteenth Australasian International Conference on Speech Science and Technology* (pp. 313–316). Parramatta, NSW: Australasian Speech Science and Technology Association.

Ching, T. Y. C., van Wanrooy, E., & Dillon, H. (2007). Binaural–bimodal fitting or bilateral implantation for managing severe to profound deafness: A review. *Trends in Amplification*, *11*, 161–192.

Cluff, M. S., & Luce, P. A. (1990). Similarity neighborhoods of spoken two-syllable words: Retroactive effects on multiple activation. *Journal of Experimental Psychology: Human Perception and Performance*, *16*, 551–563.

Cooper, N., Cutler, A., & Wales, R. (2002). Constraints of lexical stress on lexical access in English: Evidence from native and non-native listeners. *Language and Speech*, *45*, 207–228.

Cutler, A. (1980). Errors of stress and intonation. In V. A. Fromkin (Ed.), *Errors in linguistic performance: Slips of the tongue, ear, pen and hand* (pp. 67–80). New York: Academic Press.

Cutler, A. (1986). *Forbear* is a homophone: Lexical prosody does not constrain lexical access. *Language and Speech*, *29*, 201–220.

Cutler, A. (2005). Lexical stress. In D. B. Pisoni & R. E. Remez (Eds.), *The handbook of speech perception* (pp. 264–289). Oxford: Blackwell.

Cutler, A., & Bruggeman, L. (2013). Vocabulary structure and spoken-word recognition: Evidence from French reveals the source of embedding asymmetry. In F. Bimbot, C. Cerisara, C. Fougeron, et al. (Eds.), *Proceedings of the 14th Annual Conference of the International Speech Communication Association (Interspeech 2013)* (pp. 2812–2816). Lyon: International Speech Communication Association.

Cutler, A., & Carter, D. M. (1987). The predominance of strong initial syllables in the English vocabulary. *Computer Speech and Language*, *2(3–4)*, 133–142.

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., Norris, D., & Sebastián-Gallés, N. (2004). Phonemic repertoire and similarity within the vocabulary. In S. Kin & M. J. Bae (Eds.), *Proceedings of the 8th International Conference on Spoken Language Processing (Interspeech 2004-ICSLP)* (pp. 65–68). Seoul: Sunjijn.

Cutler, A., Otake, T., & Bruggeman, L. (2012). Phonologically determined asymmetries in vocabulary structure across languages. *Journal of the Acoustical Society of America*, *132*, EL155–160.

Cutler, A., & Pasveer, D. (2006). Explaining cross-linguistic differences in effects of lexical stress on spoken-word recognition. In R. Hoffmann & H. Mixdorff (Eds.), *Speech prosody 2006: Third International Conference*. Dresden: TUD Press.

Cutler, A., & van Donselaar, W. (2001). *Voornaam* is not (really) a homophone: Lexical prosody and lexical access in Dutch. *Language and Speech*, *44*, 171–195.

Cutler, A., Wales, R., Cooper, N., & Janssen, J. (2007). Dutch listeners' use of suprasegmental cues to English stress. In J. Trouvain & W. J. Barry (Eds.), *Proceedings of the 16th International Congress of Phonetics Sciences (ICPhS 2007)* (pp. 1913–1916). Dudweiler, Germany: Pirrot.

Dahan, D., Magnuson, J. S., Tanenhaus, M. K., & Hogan, E. M. (2001). Subcategorical mismatches and the time course of lexical access: Evidence for lexical competition. *Language and Cognitive Processes*, *16*, 507–534.

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

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

Dohen, M., Lœvenbruck, H., Cathiard, M.-A., & Schwartz, J.-L. (2004). Visual perception of contrastive focus in reiterant French speech. *Speech Communication*, *44(1–4)*, 155–172.

Domahs, U., Genc, S., Knaus, J., et al. (2013). Processing (un-)predictable word stress: ERP evidence from Turkish. *Language and Cognitive Processes*, *28*, 335–354.

Dorman, M. F., Gifford, R. H., Spahr, A. J., & McKarns, S. A. (2008). The benefits of combining acoustic and electric stimulation for the recognition of speech, voice and melodies. *Audiology & Neuro-Otology*, *13*, 105–112.

Dupoux, E., Pallier, C., Sebastian, N., & Mehler, J. (1997). A destressing "deafness" in French? *Journal of Memory and Language*, *26(3)*, 406–421.

Dupoux, E., & Peperkamp, S. (2002). Fossil markers of language development: Phonological "deafnesses" in adult speech processing. In J. Durand (Ed.), *Phonetics, phonology, and cognition* (pp. 168–190). Oxford: Oxford University Press.

Dupoux, E., Peperkamp, S., & Sebastián-Gallés, N. (2001). A robust method to study stress "deafness." *Journal of the Acoustical Society of America*, *110*, 1606–1618.

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

Fortescue, M., Mithun, M., & Evans, N. (Eds.). (2017). *The Oxford handbook of polysynthesis*. Oxford: Oxford University Press.

Fowler, C. A. (1995). Acoustic and kinematic correlates of contrastive stress accent in spoken English. In F. Bell-Berti & L. J. Raphael (Eds.), *Producing speech: Contemporary issues: For Katherine Safford Harris* (pp. 355–373). New York: AIP Press.

Friedrich, C. K., Kotz, S. A., Friederici, A. D., & Gunter, T. C. (2004). ERPs reflect lexical identification in word fragment priming. *Journal of Cognitive Neuroscience*, *16*, 541–552.

Fromkin, V. A. (1976). Putting the emPHAsis on the wrong sylLABle. In L. M. Hyman (Ed.), *Studies in stress and accent* (pp. 15–26). Los Angeles: Department of Linguistics, University of Southern California.

Garami, L., Ragó, A., Honbolygó, F., & Csépe, V. (2017). Lexical influence on stress processing in a fixed-stress language. *International Journal of Psychophysiology*, *117*, 10–16.

Gifford, R. H., Dorman, M. F., Brown, C. A., & Spahr, A. J. (2012). Hearing, psychophysics, and cochlear implantation: Experiences of older individuals with mild sloping to profound sensory hearing loss. *Journal of Hearing Science*, *2*, 9–17.

Holt, C. M., Demuth, K., & Yuen, I. (2016). The use of prosodic cues in sentence processing by prelingually deaf users of cochlear implants. *Ear and Hearing*, *37*, e256–262.

Holt, C. M., & McDermott, H. J. (2013). Discrimination of intonation contours by adolescents with cochlear implants. *International Journal of Audiology*, *52*, 808–815.

Jesse, A., & Massaro, D. W. (2010). The temporal distribution of information in audiovisual spoken-word identification. *Attention, Perception, & Psychophysics*, *72*, 209–225.

Jesse, A., & McQueen, J. M. (2013). Suprasegmental lexical stress cues in visual speech can guide spoken-word recognition. *Quarterly Journal of Experimental Psychology*, *67*, 793–808.

Jesse, A., Poellmann, K., & Kong, Y.-Y. (2017). English listeners use suprasegmental cues to lexical stress early during spoken-word recognition. *Journal of Speech, Language, and Hearing Research*, *60*, 190–198.

Jesse, A., Vrignaud, N., Cohen, M. A., & Massaro, D. W. (2000). The processing of information from multiple sources in simultaneous interpreting. *Interpreting*, *5*, 95–115.

Jongenburger, W. (1996). *The role of lexical stress during spoken-word processing* [PhD thesis, University of Leiden]. The Hague: Holland Academic Graphics.

Kabak, B., Maniwa, K., & Kazanina, N. (1999). Listeners use vowel harmony and word-final stress to spot nonsense words: A study of Turkish and French. *Laboratory Phonology*, *1*, 207–224.

Kelso, J. A. S., Vatikiotis-Bateson, E., Saltzman, E. L., & Kay, B. (1985). A qualitative dynamic analysis of reiterant speech production: Phase portraits, kinematics, and dynamic modeling. *Journal of the Acoustical Society of America*, *77*, 266–280.

Kong, Y.-Y., & Jesse, A. (2017). Low-frequency fine-structure cues allow for the online use of lexical stress during spoken-word recognition in spectrally degraded speech. *Journal of the Acoustical Society of America*, *141*, 373–382.

Kong, Y.-Y., Stickney, G. S., & Zeng, F.-G. (2005). Speech and melody recognition in binaurally combined acoustic and electric hearing. *Journal of the Acoustical Society of America*, *117*, 1351–1361.

Koster, M., & Cutler, A. (1997). Segmental and suprasegmental contributions to spoken-word recognition in Dutch. In *Proceedings of EUROSPEECH 97* (pp. 2167–2170). Rhodes, Greece: International Speech Communication Association.

Lehiste, I., & Peterson, G. (1959). Vowel amplitude and phonemic stress in American English. *Journal of the Acoustical Society of America*, *31*, 428–35.

MacLeod, A., & Summerfield, Q. (1987). Quantifying the contribution of vision to speech perception in noise. *British Journal of Audiology*, *21*, 131–141.

Magnuson, J. S., Dixon, J. A., Tanenhaus, M. K., & Aslin, R. N. (2007). The dynamics of lexical competition during spoken word recognition. *Cognitive Science*, *31*, 133–156.

Marslen-Wilson, W. (1990). Activation, competition, and frequency in lexical access. In G. T. M. Altmann (Ed.), *Cognitive models of speech processing: Psycholinguistic and computational perspectives* (pp. 148–172). Cambridge, MA: MIT Press.

Mattys, S. L. (2000). The perception of primary and secondary stress in English. *Perception & Psychophysics*, *62*, 253–265.

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

McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2002). Gradient effects of within-category phonetic variation on lexical access. *Cognition*, *86*, B33–42.

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*, 621–638.

McQueen, J. M., & Viebahn, M. C. (2007). Tracking recognition of spoken words by tracking looks to printed words. *Quarterly Journal of Experimental Psychology*, *60*, 661–671.

Mehler, J., Peña, M., Nespor, M., & Bonatti, L. (2006). The "soul" of language does not use statistics: Reflections on vowels and consonants. *Cortex*, *42*, 846–854.

Meister, H., Landwehr, M., Pyschny, V., et al. (2009). The perception of prosody and speaker gender in normal-hearing listeners and cochlear implant recipients. *International Journal of Audiology*, *48*, 38–48.

Moore-Cantwell, C., & Sanders, L. (2017). Effects of probabilistic phonology on the perception of words and nonwords. *Language, Cognition and Neuroscience*, *33*, 148–164.

Morris, D., Magnusson, L., Faulkner, A., et al. (2013). Identification of vowel length, word stress, and compound words and phrases by postlingually deafened cochlear implant listeners. *Journal of the American Academy of Audiology*, *24*, 879–890.

Munhall, K. G., Jones, J. A., Callan, D. E., et al. (2004). Visual prosody and speech intelligibility: Head movement improves auditory speech perception. *Psychological Science*, *15*, 133–137.

Nazzi, T., & Cutler, A. (2019). How consonants and vowels shape spoken-language recognition. *Annual Review of Linguistics*, *5*, 25–47.

Ortega-Llebaria, M., Gu, H., & Fan, J. (2013). English speakers' perception of Spanish lexical stress: Context-driven L2 stress perception. *Journal of Phonetics*, *41*, 186–197.

Peng, S. C., Lu, N., & Chatterjee, M. (2009). Effects of cooperating and conflicting cues on speech intonation recognition by cochlear implant users and normal hearing listeners. *Audiology & Neuro-Otology*, *14*, 327–337.

Peperkamp, S., Vendelin, I., & Dupoux, E. (2010). Perception of predictable stress: A cross-linguistic investigation. *Journal of Phonetics*, *38*, 422–430.

Potisuk, S., Gandour, J., & Harper, M. P. (1996). Acoustic correlates of stress in Thai. *Phonetica*, *53*, 200–220.

Reinisch, E., Jesse, A., & McQueen, J. M. (2010). Early use of phonetic information in spoken word recognition: Lexical stress drives eye movements immediately. *Quarterly Journal of Experimental Psychology*, *63*, 772–783.

Risberg, A., & Lubker, J. (1978). Prosody and speechreading. *Speech Transmission Laboratory Quarterly Progress Status Report*, *4*, 1–16.

Scarborough, R., Keating, P., Mattys, S. L., et al. (2009). Optical phonetics and visual perception of lexical and phrasal stress in English. *Language and Speech*, *52*, 135–175.

Slowiaczek, L. M. (1990). Effects of lexical stress in auditory word recognition. *Language and Speech*, *33*, 47–68.

Slowiaczek, L. M. (1991). Stress and context in auditory word recognition. *Journal of Psycholinguistic Research*, *20*, 465–481.

Small, L. H., Simon, S. D., & Goldberg, J. S. (1988). Lexical stress and lexical access: Homographs versus nonhomographs. *Perception & Psychophysics*, *44*, 272–280.

Smith, Z. M., Delgutte, B., & Oxenham, A. J. (2002). Chimaeric sounds reveal dichotomies in auditory perception. *Nature*, *416*, 87–90.

Söderström, P., Horne, M., Frid, J., & Roll, M. (2016). Pre-activation negativity (PrAN) in brain potentials to unfolding words. *Frontiers in Human Neuroscience*, *10*, art. 512.

Soto-Faraco, S., Sebastián-Gallés, N., & Cutler, A. (2001). Segmental and suprasegmental mismatch in lexical access. *Journal of Memory and Language*, *45(3)*, 412–432.

Spitzer, S., Liss, J., Spahr, T., Dorman, M., & Lansford, K. (2009). The use of fundamental frequency for lexical segmentation in listeners with cochlear implants. *Journal of the Acoustical Society of America*, *125*, EL236–241.

Srinivasan, R. J., & Massaro, D. W. (2003). Perceiving prosody from the face and voice: Distinguishing statements from echoic questions in English. *Language and Speech*, *46*, 1–22.

Sulpizio, S., & McQueen, J. M. (2012). Italians use abstract knowledge about lexical stress during spoken-word recognition. *Journal of Memory and Language*, *66(1)*, 177–193.

Sumby, W. H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society of America*, *26*, 212–215.

Suomi, K., Toivanen, J., & Ylitalo, R. (2003). Durational and tonal correlates of accent in Finnish. *Journal of Phonetics*, *31*, 113–138.

Tagliapietra, L., & Tabossi, P. (2005). Lexical stress effects in Italian spoken word recognition. In B. G. Bara, L. Barsalou, & M. Bucciarelli (Eds.), *Proceedings of the XXVII Annual Conference of the Cognitive Science Society* (pp. 2140–2144). Stresa, Italy: Lawrence Erlbaum.

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

van Donselaar, W., Koster, M., & Cutler, A. (2005). Exploring the role of lexical stress in lexical recognition. *Quarterly Journal of Experimental Psychology Section A*, *58*, 251–273.

van Heuven, V. J. (1988). Effects of stress and accent on the human recognition of word fragments in spoken context: Gating and shadowing. In W. A. Ainsworth & J. N. Holmes (Eds.), *Proceedings of Speech '88, the 7th FASE Symposium* (pp. 811–818). Edinburgh: Institute of Acoustics.

van Leyden, K., & van Heuven, V. J. (1996). Lexical stress and spoken word recognition: Dutch vs. English. In C. Cremers & M. den Dikken (Eds.), *Linguistics in the Netherlands 1996* (Vol. *13*, pp. 159–170). Amsterdam: John Benjamins.

van Ooijen, B. (1996). Vowel mutability and lexical selection in English: Evidence from a word reconstruction task. *Memory & Cognition*, *24*, 573–583.

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.

Wagner, A. (2013). Cross-language similarities and differences in the uptake of place information. *Journal of the Acoustical Society of America*, *133*, 4256–4267.

Wagner, A., Ernestus, M., & Cutler, A. (2006). Formant transitions in fricative identification: The role of native fricative inventory. *Journal of the Acoustical Society of America*, *120*, 2267–2277.

Warner, N. L., Hernandez, G. G., Park, S., & McQueen, J. M. (2018). A replication of competition and prosodic effects on spoken word recognition. *Journal of the Acoustical Society of America*, *143*, 1921.

White, K. S., Chambers, K. E., Miller, Z., & Jethava, V. (2018). Listeners learn phonotactic patterns conditioned on suprasegmental cues. *Quarterly Journal of Experimental Psychology*, *70*, 2560–2576.

Williams, B. (1985). Pitch and duration in Welsh stress perception: The implications for intonation. *Journal of Phonetics*, *13*, 381–406.

Woodson, E. A., Reiss, L. A. J., Turner, C. W., et al. (2010). The Hybrid cochlear implant: A review. *Advances in Oto-Rhino-Laryngology*, *67*, 125–134.

Yu, J., Mailhammer, R. & Cutler, A. (2020). Vocabulary structure affects word recognition; Evidence from German listeners. In In N. Minematsu, M. Kondo,

T. Arai, & R. Hayashi (Eds.), *Proceedings of Speech Prosody 2020* (pp. 474-478). Tokyo: ISCA.

Zeng, F.-G., Nie, K., Stickney, G. S., et al. (2005). Speech recognition with amplitude and frequency modulations. *Proceedings of the National Academy of Sciences of the United States of America*, *102*, 2293–2298.

Zhang, T., Dorman, M. F., & Spahr, A. J. (2010). Information from the voice fundamental frequency (F0) region accounts for the majority of the benefit when acoustic stimulation is added to electric stimulation. *Ear and Hearing*, *31*, 63–69.

Zora, H., Heldner, M., & Schwarz, I.-C. (2016). Perceptual correlates of Turkish word stress and their contribution to automatic lexical access: Evidence from early ERP components. *Frontiers in Neuroscience*, *10*, 7.