IN THRALL TO THE VOCABULARY

Anne Cutler

The MARCS Institute, University of Western Sydney, Sydney, Australia
a.cutler@uws.edu.au

Vocabularies contain hundreds of thousands of words built from only a handful of phonemes; longer words inevitably tend to contain shorter ones. Recognising speech thus requires distinguishing intended words from accidentally present ones. Acoustic information in speech is used wherever it contributes significantly to this process; but as this review shows, its contribution differs across languages, with the consequences of this including: identical and equivalently present information distinguishing the same phonemes being used in Polish but not in German, or in English but not in Italian; identical stress cues being used in Dutch but not in English; expectations about likely embedding patterns differing across English, French, Japanese.

THE *MESS-IN-THE-MESSAGE* PROBLEM

At St John's Church in Darlinghurst, Sydney, a sign reads "Only God can turn a MESS into a message, a TEST into a testimony, a TRIAL into a triumph, a VICTIM into a victory". Full marks to the writer for effective use of the resources that English offers! But the proposition of exclusivity is, in fact, inaccurate, because turning a *mess* into a *message*, and the like, is what every speaker of English has to do every day to understand speech. The vocabulary gives us no choice in this matter. By the standards of the world's languages, English has many phonemes (42 in the Australian variety [1], where the world cross-language mean is 31 and the mode 25 [2]; Figure 1). If distinctiveness were a priority, 42 is still a tiny number from which to construct the hundreds of thousands of words of the English vocabulary. All vocabularies are like this: a huge number of words, constructed from a trivially small number of contrastive speech sounds. The inevitable result is that there is very little distinctiveness. Words closely resemble other words (*word: bird, curd, herd, weed, wide, wade, work, whirl, worse*...). Further, short words occur accidentally in longer ones (*test* in *testimony*, etc.); and many longer words, not at all related, begin in the same way (*trial* and *triumph*, etc.).
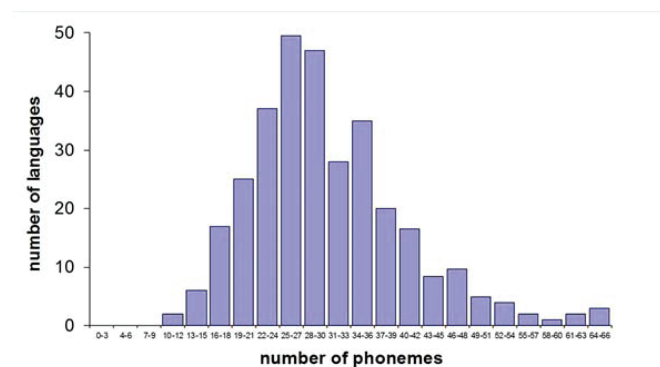

Figure 1. Phoneme inventory size across languages. In a representative sample of world languages, the count (vertical axis) of languages by size of phoneme inventory (in increments of three, the smallest set being 10-12 phonemes, the largest 64-66). The mode is 25; Australian English (42 phonemes) is in the upper tail of the distribution.

Spoken language thus presents more word recognition options than is desirable from the listener's perspective; not only the string of words intended by a speaker, but accidental embedded words within some of those intended words, and beyond that, words embedded across the intended words as well (*victim* in *evict immediately*, *worse* in *were stopping*, etc.). Of course many of these options will completely mismatch the context and can easily be rejected if recognised; but it is rare for listeners to become at all aware of the multiple alternative options, because the efficiency of word recognition is such that the intended string is usually settled upon rapidly and the unintended other options are efficiently discarded. Though many possibilities are briefly available, and indeed compete among one another for recognition, any option that the unfolding speech input mismatches can be immediately discarded.

With experimental methods from the psycholinguistic laboratory it is possible to discern traces of the fleeting presence of rejected competitor words. One such method (eyetracking) offers visual representations of word options. For example, listeners who hear *Now pick up the sandal....* while looking at a display showing a sandal, a sandwich, and two other objects, typically look to each of the objects with names that begin *sand-* until the moment at which the speech makes clear that it is one and not the other. The point here is that no one waits until a whole word has been presented; the input is assessed continuously for evidence of what it might be. Interestingly, if one of the other objects is, say, a candle (a name differing from the target word only in initial sound), that too attracts looks – not as many as the sandwich, but more than some object with a dissimilar name [3]. (There is more on why this might be useful in section 4 "Test in Testimony Versus Detest" below.)

In another method (cross-modal fragment priming) listeners decide whether a written string of letters is a real word or not, as they listen to some speech. The critical question is how the speech input affects availability of word options, with the availability revealed by how fast a string can be acknowledged as indeed a real word. Typically, response to the same word is compared in three situations: (a) when the spoken input is a completely different word or fragment (this is a baseline for

how easy the word is to recognise); (b) when the input is all or part of the same word (this produces priming – the word is heard and seen at the same time, so this should make it highly available); and (c) when the speech partly overlaps with, but mismatches the written word [4].

For example, listeners might respond to DELIVER, while hearing a word fragment ending a neutral sentence such as *The password for this week is...* The baseline control fragment could be *supplem-*, while the matching fragment is *deliv-*. Positive lexical decision responses to DELIVER will be quite a lot faster in the latter case. Then suppose the case (c) fragment is *delish-* (from the word *delicious*). What we see then is that responses to DELIVER are slower than after the baseline fragment (*supplem-*), i.e., the word form *deliver* is less available than even in that neutral situation. Its availability has actually been inhibited. The first few phonemes of the target word made both *deliver* and *delicious* available, but on arrival of the [ʃ] that matched *delicious* but mismatched *deliver*, the mismatched word was rejected. The inhibition of responses to written DELIVER after spoken *delish* reveals auditory *deliver's* temporary, ultimately unsuccessful, presence.

Exactly the same happens when we hear *trial* or *triumph*, *victim* or *victory*; at some point, the input forces us to choose one option over the other. Embedded words such as *mess* in *message* may be fleetingly the strongest available candidate, but they too are eventually overridden by stronger evidence for a longer candidate word. And given the structure of all vocabularies, this type of efficient continuous evaluation of the incoming acoustic signal entails constant resolution of *mess-in-the-message* issues, whenever we hear speech.

SELECTIVE LISTENING: PHONEMES

Speech signals are multidimensional, and listeners can call on several information sources to make effectively the same decision. Decisions involve choice between word options, and hence between speech sounds that crucially distinguish them. Deciding that one is hearing *victim* or *victory* becomes possible when the sixth phoneme turns out to be [m] or [r]. Phonemes are famously not separated entities in speech signals, but overlap, forming the "speech code" [5]. Some sounds can not be reliably identified without knowing what the following sound is, and upcoming sounds may be cued by signals that are present in the speech before the primary speech gestures for the upcoming sound have actually been made. Both types of cue, direct and contextual, are available for listeners to use, and use them they do. The fifth phoneme in *victim* and *victory* is a reduced vowel that will effectively signal already whether an [m] or [r] follows, since both a following consonant's place and manner of articulation (e.g., bilabial and nasal in [m], alveolar and approximant in [r]) have detectable effect in a preceding vowel (similarly, a vowel influences a preceding consonant too).

In some cases, however, available information is just ignored. Whether or not coarticulatory cues are used can be tested by cross-splicing speech signals. Using this method, Harris [6] discovered that listeners use the transitional information from [f] to an immediately following vowel to decide that they are hearing [f]; [f] from *fi* spliced to [i] from

si, *shi* or *thi* was hard to recognise. This showed that the transitional information was being used to inform decisions. But the results for [s] were quite different: [s] from *si* was as easy to recognize whether the vowel came from *si*, *shi*, *thi* or *fi*. Harris' interpretation invoked the acoustic signal, which gives clearer information for the [s] of *sea* than for the [f] of *fee*. But it turned out (many years later) that the underlying reason was not acoustics, but the vocabulary! The acoustics of [s] and [f] are the same in Dutch, German, and Italian as in English, but Harris' result does not replicate in those languages: Dutch, German, and Italian listeners all ignore transitional information for both [f] and [s] [7, 8]. In contrast, listeners from Madrid (speakers of Castilian Spanish) exhibit the same asymmetry as English-speakers, using transition information for [f] but not for [s].

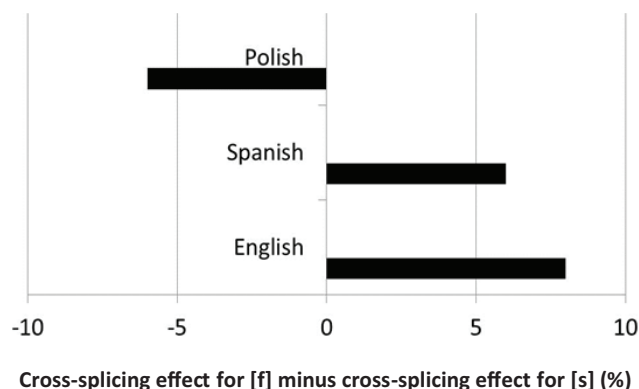Cross-splicing effect for [f] minus cross-splicing effect for [s] (%)

Figure 2. The cross-splicing effect for [f] (how much longer it takes to identify a token of [f] that has been spliced onto a vowel originally uttered after a different sound, vs. after an [f]) minus the same cross-splicing effect for [s], in English and Spanish (showing a positive difference, i.e., a greater effect for [f]) versus Polish (negative, i.e., greater for [s]) [7]. In German and Dutch, the same experiment showed virtually no effect for either sound.

What do speakers of English and of Castilian Spanish have that speakers of Dutch, German and Italian don't? Wagner [7, 8] found the answer: both English and Castilian Spanish have the sound [θ], the dental fricative as in *thick* and *thin* and Castilian *gracias*. This sound is notoriously difficult to distinguish from [f] [9]; so if your vocabulary's phoneme set contains both sounds, it pays to attend to all the information there is in identifying them. Dutch, German and Italian have no spectrally similar fricatives at all, and English and Spanish have no sibilants that could be easily confused with [s]. That is why the attention to transitions occurs, in these experiments, only for [f] and only in English and Spanish. Suppose, though, that your language did have spectrally similar sibilants? Then it would pay to attend to the transitional information for [s]. Polish is such a language, and Polish listeners in Wagner's study [7] indeed showed the reverse pattern to English and Spanish listeners – for them, cross-spliced [s] was harder than cross-spliced [f]. They used the transitional information to identify [s] but not [f] (see Figure 2). This was the crucial sign that the results were due to the language-specific phoneme sets from which the vocabularies were built.

Sorting out which words are really present in a speech stream, and which word forms are only accidentally or partially present, requires listeners to identify as rapidly as possible the speech sounds (phonemes) being uttered. In some languages, pairs of quite similar phonemes are best distinguished by attending to their transition into and out of abutting phonemes. Such transitional information is always there in the acoustic signal, but in some other languages there is no need to take notice of it.

SELECTIVE LISTENING: STRESS

Not only segments make distinctions between words. Some languages call on an extra dimension for this task. In tone languages (e.g., Mandarin), pitch movements in syllables distinguish words. And in lexical-stress languages (e.g., English or Spanish), words can differ in stress alone: *insight* is stressed on the first syllable, *incite* on the second, though the phonemes in each word are the same.
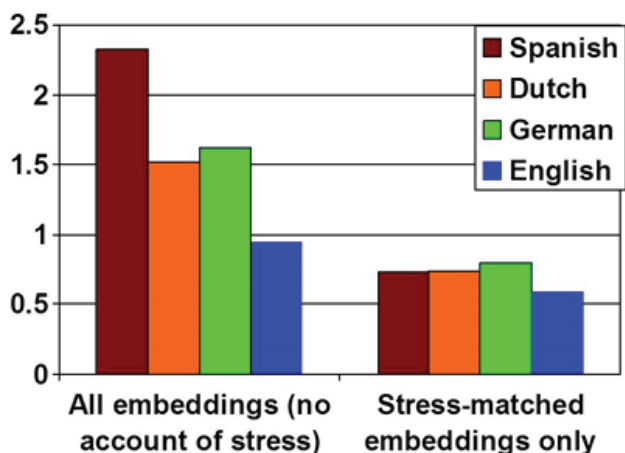

Figure 3. The effect of taking stress into account in computing embedding frequency across the vocabulary, for four languages. If stress is not considered, English *enterprise* contains *enter* and *prize*, and *settee* has *set* and *tea*. If primary stress must match, *enterprise* has only *enter*, and *settee* only *tea*. This stress reduction removes about two-thirds of embedded words from the calculation for Spanish, and about half for Dutch or German. Per word of speech each language then ends up with, on average, less than one embedding. English averaged less than one embedding without considering stress.

Such an extra dimension is particularly useful with smaller phoneme repertoires – e.g., Mandarin (for tone) or Spanish (for stress), each of which have 25 phonemes (world-wide the most common number). The Spanish *mess-in-message* problem (the average number of embeddings in real speech[1]) is reduced by 68.5% if it is computed taking stress pattern as well as segments into consideration [10]; on average 2.32 competitor words per really uttered word if stress pattern is not considered, but only 0.73 competitors per real word once stress is controlled. Obviously, it rewards Spanish listeners to take account of stress as they listen to speech. Indeed they do so. In cross-modal priming experiments in Spanish [4], a mismatching spoken fragment such as *prinCI-* (from *prinCIpio*

'principle'; the second syllable has primary stress) paired with written PRINCIPE (*PRINcipe*, 'prince') inhibited responses in just the same way as a mismatching segment did (e.g., *histe-* from *histeria* 'hysteria' with HISTORIA 'history').

Neither the competitor reduction in Spanish, nor the Spanish cross-modal fragment priming findings, replicate however in English [10, 11]. This is due not only to English's many phonemes, but also to the way English deals with stress. One of the phonemes that Spanish does not have is the reduced and central vowel schwa [ə]. English has it, though, in a majority of unstressed syllables. So, minimal stress pairs (e.g., *insight-incite*) are rare in English. Most English word pairs that are spelled the same way but differ in stress have spectrally different vowels: e.g., the *REcord* vs. to *reCORD*. The stressed first syllable always has a full vowel, but the unstressed first syllable contains schwa.

All listeners must process segments, of course. For listeners to English, just attending to segments gets most stress differences too. There is little more yield from attending also to suprasegmental cues to stress – the differences of duration, amplitude and pitch that Spanish listeners use to distinguish the initial syllables of *principe* versus *principio*. Note that these differences are indeed there in English speech whenever we compare English words with the same syllable contrasting primary versus secondary stress – e.g., the initial syllables of *MUsic/muSEum*, or *ADmiral/admiRAtion*. Table 1 shows acoustic measures from a set of such pairs.

Table 1. Mean acoustic measures (of duration, F0, rms amplitude and spectral tilt) across a female speaker of Australian English's utterance of 21 word pairs with the same first syllable, differently stressed (e.g. *mu-* with primary stress from *MUsic*, vs. *mu-* with secondary stress from *muSEum*) [14].

	Primary	Secondary
duration	381 ms	350 ms
mean F0	208 Hz	180 Hz
max F0	224 Hz	202 Hz
sd F0	12.9 Hz	11.3 Hz
mean rms	641	511
sd rms	229	174
spectral tilt	.909	.202

However, pairs like that are rare in English. So, calculation of competitor reduction with versus without stress information reveals a much smaller reduction in English than in Spanish. Importantly, the average number of competitors per real word is already below one (0.94) even without considering stress [10] (see Figure 3). One would not be surprised to find English listeners failing to attend to the suprasegmental information, and indeed that is exactly what cross-modal studies revealed: there is no significant inhibition for stress-mismatching fragments in English (e.g. [11]; these studies used matched-vowel fragments such as the first two syllables of *ADmiral/admiRAtion*). Figure 4 shows the relative amount of inhibition across languages.
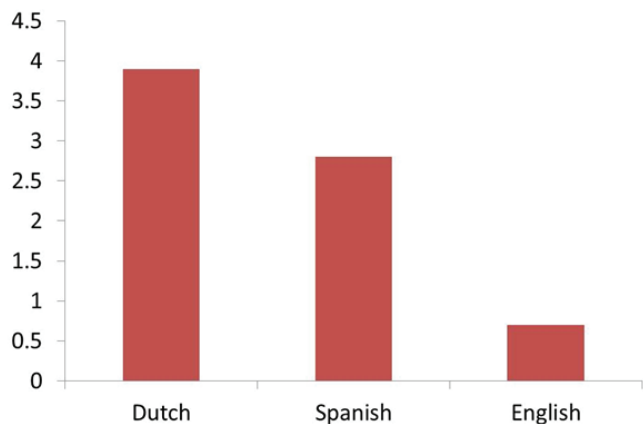
Figure 4. How much inhibition does stress mismatch cause? Across three languages, responses to a (written) word were compared after a stress-mismatching spoken fragment versus after an unrelated control fragment. A mismatch example in Dutch is DOMINANT 'dominant' after hearing *domi-* from *DOminee* 'pastor'; in Spanish, COMEDOR 'dining- room' after *come-* from *coMEdia* 'comedy'; in English, ENTERTAIN after *enter-* from *ENterprise*. In each of these cases the written target word would be stressed on the third syllable (N.B. the effect is equivalent wherever a stress mismatch occurs). The percentage to which responses are slower after stress mismatch than after a control (e.g. *pano-*) is highly significant in Dutch and Spanish, but not in English (where there is not even one percentage point difference).

Stress languages such as Dutch and German, though they are closely related to English with a very similar phonology of stress, differ from English in that they have a much lesser tendency to use the vowel schwa in unstressed syllables. In consequence, considering stress in computing competitor numbers has a larger effect in each of these languages [12], so that listeners benefit (more than English listeners do) from computing stress in recognising words. Dutch listeners, indeed, show the significant inhibition from stress mismatch in the crossmodal task [13], inter alia with primary versus secondary-stress contrasts as in *DOminee* 'pastor' vs. *domiNANT*, 'dominant'.

A useful side effect for listeners of these languages occurs, then, when they use English as a second language. As Table 1 shows, the suprasegmental cues to stress are fully there in English speech, even if English listeners ignore them when recognising words like *admiral* and *admiration*; but Dutch listeners, given English words, do <u>not</u> ignore this information, and actually outdo native English-speakers in correctly assigning syllables differing in stress [11, 14, 15]. Of course, this does not yield them a great benefit in competition reduction, but it may compensate for other ways in which listening to a second language is harder than listening to the native language.

With stress distinctions between words, as with segmental distinctions, acoustic information that is present in speech may or may not be exploited in the recognition of spoken words; the vocabulary dictates how much use is made of it.

TEST IN TESTIMONY VERSUS DETEST

Short words can be embedded anywhere in a longer carrier word (there is a *mess* in *domestic* as well as in *message*, and a *test* not only in *testimony* but also in *intestine* and *detest*). These competitors do not all carry equal weight in the recognition process. One of the first findings to appear in studies of embedding was that there is more embedding at the beginning of carrier words than at the end [16] – and English and Spanish both showed the same pattern in this respect [10]. This asymmetry is important, because a speech signal remains ambiguous as long as words overlap from their onsets. Thus, as noted, onset competitors such as the sandwich (given spoken *sandal*) attract more looks in eyetracking studies than offset competitors such as the candle [3]. Interestingly, though, this pattern appears to be be under listener control; in a little background noise – even when the target word itself is not masked by the noise at all – offset competitors are no longer so disadvantaged [17]. It seems that listeners allow for the possibility that their perception might be impaired in difficult listening conditions, and adjust their expectations about word similarity to deal with this, considering other words ending in the same way as the apparent input as well as words beginning in the same way.
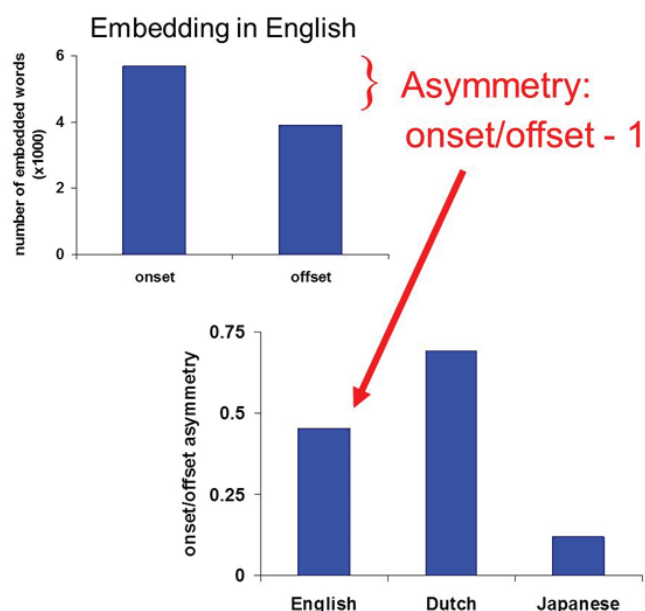

Figure 5. In English, a 1.45:1 ratio holds for the number of embedded words at a carrier word onset (e.g., *enter* in *enterprise*) versus at offset (e.g., *prize* in *enterprise*). The asymmetry (.45) is plotted in the lower graph, where English is compared with Dutch (more asymmetry) and Japanese (virtually no asymmetry).

Because of the importance of competition in understanding speech recognition, asymmetry in the distribution of competitors in the vocabulary will impact upon listening. But once more, not for all listeners equivalently – because this pattern too varies across vocabularies. It is found in English, indeed in all the Germanic languages, and in Spanish; but it is not found in the vocabulary of Japanese. There, due to Japanese

word structure, the amount of embedding at carrier word onset and offset is almost the same, as Figure 5 shows [18].

What really causes such asymmetry? As noted, the Germanic languages all have lexical stress, as does Spanish. Does this create asymmetry in embedding patterns? The Germanic languages in particular all exhibit a strong tendency towards initial stress [19]. Combine this with the tendency (strongest of all in English) for unstressed syllables to be weak, and it is clear that syllables at the ends of words are less likely to happen to correspond to other stand-alone words in the vocabulary. (The asymmetry is most marked in German, as German tends to add an unstressed syllable, pronounced as schwa, to words with mono- or disyllabic cognates in English or Dutch: *cat*, *kat*, *Katze* in English, Dutch, German, respectively; *pill*, *pil*, *Pille*; *cigar*, *sigaar*, *Zigarre*; *guitar*, *gitaar*, *Gitarre*; etc.).

Japanese has neither lexical stress nor the vowel schwa. But Japanese also does not have suffixing morphology, which is another prime candidate for source of the embedding asymmetry. Suffixes, either inflections (*talking*, *spotted*, *boxes*) or derivations (*business*, *vibration*, *government*) also tend to be weak syllables that cannot stand alone. All of the stress languages with the embedding asymmetry also have suffixes.

Either stress or suffixes or both could underlie the observed asymmetry. The best way to sort out this question is to find a language which has one of these structural features, but not the other. French is such a language: It has suffixes, but it does not have lexical stress. If only stress is necessary, and sufficient, for an embedding asymmetry, French will pattern like Japanese and show no asymmetry; if suffixing is necessary and sufficient, then French will also show an asymmetry – it should pattern like English. If both factors contribute, we might expect it to fall roughly halfway between the effect found in English and in Japanese respectively.
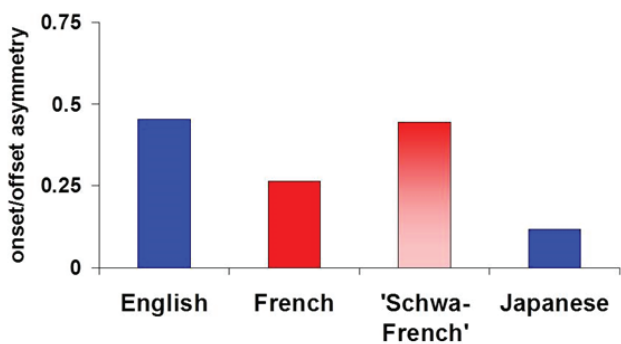

Figure 6. The same onset/offset asymmetry as shown for English and Japanese in Figure 5, now also for French (like English, with suffixes, but like Japanese, without weak final syllables) and for an imaginary variety of French, "schwa- French", in which weak final syllables containing schwa are added as well.

Using lexical resources for French [20], and computing embedding patterns in its vocabulary as in the other languages, produces a pattern that is indeed halfway between those of English and of Japanese [21]: see Figure 6. Suffixes thus encourage embedding asymmetry. Stress may add further to this effect. The role of stress, we argued, is played out mainly in the placement of weak syllables, i.e., those with schwa. French, though it does not have lexical stress, does have the vowel schwa in its phoneme repertoire. Moreover, it is phonologically legal to expand the role of schwa in a way that would make French more Germanic-like; in certain French dialects, words such as *ville*, *petite*, *bonne*, spelled with a final letter *e* which in standard French is silent, have an extra syllable pronounced as schwa [22]. If all such words in the lexicon of standard French are assumed to be indeed given this extra final schwa in their pronunciation, and the embedding patterns across the vocabulary are then recalculated in the same way, the embedding asymmetry in this artificial variety of French becomes almost exactly that of English [21]. This is again shown in Figure 6.

Thus, both a phonological difference between stressed and unstressed syllables, and the presence of morphological affixes, affect the distribution of embedded words in the vocabulary and hence the patterns of competition affecting speech presented to listeners. English, with both stress and suffixes, has many more words with initial embeddings (*testimony*) than final embeddings (*detest*). This causes patterns of listening that differ from those in languages such as Japanese, which has neither of these precipitating factors, and in consequence, no such embedding asymmetry.

CONCLUSION

Speech acoustics presents a vast array of useful information, but some of it, sometimes, in some languages, gets ignored (even though exactly the same information is put to effective use in other languages). Information is only used where it makes a measurable difference in distinguishing one phoneme from another and hence one word from another, as well as really uttered words from words that are only accidentally present in speech.

Possibly it would be useful to master the exploitation of additional speech cues, for instance when we try to learn, as adults, a language in which those cues are used (Spanish, Dutch or Polish, to use some of the above examples). Research has not yet addressed whether appropriate training could bring this about. Indeed, we do not yet even know when child learners begin to use cues appropriately, nor whether early-bilingual users of two languages that encourage different cue use strategies switch such strategies as they switch language perception. But in general the picture for most of us is that there is nothing much we can do about which cues we use. Our native vocabulary has us in its thrall.

FOOTNOTE

[1] This is computed by calculating the embedded words in every word in the vocabulary (whereby syllable boundary match is required – so, *can* is counted in *candle* but not in *scandal* or *cant*, etc.) and then multiplying every carrier word's total of embeddings by the carrier word's frequency in a relevant corpus. This procedure thus accounts for a frequent word (e.g., *dinner*, with embedded *din*) contributing more to average listeners' experience of lexical competition than an infrequent word (*redintegrate*, with embedded *din*, *great* etc.).

REFERENCES

[1] F. Cox and S. Palethorpe, "Australian English", *Journal of the International Phonetic Association* **37**, 341-350 (2007)

[2] I. Maddieson, I. *Patterns of sounds*, Cambridge University Press, Cambridge 1984

[3] P. Allopenna, J. Magnuson and M. Tanenhaus, "Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping model", *Journal of Memory and Language* **38**, 419–439 (1998)

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

[5] A.M. Liberman, F.S. Cooper, D.P. Shankweiler and M. Studdert-Kennedy, "Perception of the speech code", *Psychological Review* **74**, 431–461 (1967)

[6] K.S. Harris, "Cues for the discrimination of American English fricatives in spoken syllables", *Language and Speech* **1**, 1–17 (1958)

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

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

[9] G.A. Miller and P. Nicely, "An analysis of perceptual confusions among some English consonants", *Journal of the Acoustical Society of America* **27**, 338-352 (1955)

[10] A. Cutler, D. Norris and N. Sebastián-Gallés, "Phonemic repertoire and similarity within the vocabulary", *Proceedings of the 8th International Conference on Spoken Language Processing*, Jeju, Korea, 2004, pp. 65–68

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

[12] A. Cutler and D. Pasveer, "Explaining crosslinguistic differences in effects of lexical stress on spoken-word recognition", *Proceedings of the 3rd International Conference on Speech Prosody*, Dresden, Germany, 2006, pp. 237–240

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

[14] A. Cutler, R. Wales, N. Cooper and J. Janssen, "Dutch listeners' use of suprasegmental cues to English stress", *Proceedings of the 16th International Congress of Phonetic Sciences*, Saarbrücken, Germany, 2007, pp. 1913–1916

[15] A. Cutler, "Greater sensitivity to prosodic goodness in non-native than in native listeners", *Journal of the Acoustical Society of America* **125**, 3522–3525 (2009)

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

[17] J.M. McQueen and F. Huettig, "Changing only the probability that spoken words will be distorted changes how they are recognized", *Journal of the Acoustical Society of America* **131**, 509-517 (2012)

[18] A. Cutler, T. Otake and L. Bruggeman, "Phonologically determined asymmetries in vocabulary structure across languages", *Journal of the Acoustical Society of America* **132**, EL155-EL160 (2012)

[19] M. Trommelen and W. Zonneveld "Word stress in the West Germanic languages", in *Word prosodic systems in the languages of Europe* ed. H. van der Hulst, (Mouton de Gruyter, Berlin, 1999) pp. 478-515

[20] B. New, C. Pallier, M. Brysbaert and L. Ferrand, "Lexique 2: A new French lexical database", *Behavior Research Methods* **36**, 516-524 (2004)

[21] A. Cutler and L. Bruggeman, "Vocabulary structure and spoken-word recognition: Evidence from French reveals the source of embedding asymmetry", *Proceedings of INTERSPEECH 2013*, Lyon, France, 2013, pp. 2812-2816

[22] J. Durand, C. Slater and H. Wise, "Observations on schwa in southern French," *Linguistics* **25**, 983-1004 (1987)