Original Paper

Phonetica 2017;74:81–106
DOI: 10.1159/000447428

Received: September 17, 2015
Accepted: June 4, 2016
Published online: October 7, 2016

Stress Effects in Vowel Perception as a Function of Language-Specific Vocabulary Patterns

Natasha Warner[a] Anne Cutler[b, c]

[a]Department of Linguistics, University of Arizona, Tucson, Ariz., USA; [b]MARCS Institute, University of Western Sydney, Sydney, N.S.W., Australia; [c]Max Planck Institute of Psycholinguistics, Nijmegen, The Netherlands

Abstract

Background/Aims: Evidence from spoken word recognition suggests that for English listeners, distinguishing full versus reduced vowels is important, but discerning stress differences involving the same full vowel (as in *mu-* from *music* or *museum*) is not. In Dutch, in contrast, the latter distinction is important. This difference arises from the relative frequency of unstressed full vowels in the two vocabularies. The goal of this paper is to determine how this difference in the lexicon influences the perception of stressed versus unstressed vowels. ***Methods:*** All possible sequences of two segments (diphones) in Dutch and in English were presented to native listeners in gated fragments. We recorded identification performance over time throughout the speech signal. The data were here analysed specifically for patterns in perception of stressed versus unstressed vowels. ***Results:*** The data reveal significantly larger stress effects (whereby unstressed vowels are harder to identify than stressed vowels) in English than in Dutch. Both language-specific and shared patterns appear regarding which vowels show stress effects. ***Conclusion:*** We explain the larger stress effect in English as reflecting the processing demands caused by the difference in use of unstressed vowels in the lexicon. The larger stress effect in English is due to relative inexperience with processing unstressed full vowels.

Introduction

Incite differs from *inside* only in a single dimension: the voicing of one phoneme. *Incite* and *insight* also differ in only one dimension: where primary stress falls. Speech has multiple levels on which contrastive information may be encoded. In all languages, minimal contrasts between segments – vowels or consonants – distinguish one word from another. In many languages, minimal contrasts in suprasegmental dimensions can also be lexically contrastive (a contrast is *suprasegmental* when a contrasting pair of

KARGER

© 2016 S. Karger AG, Basel

E-Mail karger@karger.com
www.karger.com/pho

Dr. Natasha Warner
Department of Linguistics, University of Arizona
Box 210028
Tucson, AZ 85721-0028 (USA)
E-Mail nwarner@email.arizona.edu

sequences maintains identity at the segmental level though differing in duration, pitch, etc.; Lehiste, 1970). Lexical tone languages contrast words by pitch realised on syllables, quantity languages contrast long versus short versions of segments, and lexical stress languages such as English can contrast words in the placement of prosodic salience within the word[1].

Stress languages differ in whether stress is realised solely in the suprasegmental dimensions of pitch, duration, and amplitude, or interacts to some extent with the segmental level. In Spanish, for instance, there is no such interaction, so the 5 vowels of the language can freely occur in stressed or in unstressed syllables. In English, though, stress interacts strongly with segments. English has vowel reduction, and the distinction between full and reduced vowels plays a central role in stress patterning. Two generalisations hold: a reduced vowel cannot bear stress, and all stressed syllables must contain a full vowel[2].

In principle, words in any lexical stress language can contrast solely in stress, and this includes English (because those two generalisations for English do not prohibit full vowels from appearing in unstressed syllables). However, pairs such as *INsight-inCITE*[3] are in fact rare; besides *INsight-inCITE*, English has *TRUSty-trusTEE*, *DIScount-disCOUNT*, *FOREbear-forBEAR*, and a handful more. This rarity alone suggests that the suprasegmental stress information may have little relevance in the recognition of English, and indeed, evidence from a range of empirical paradigms suggests that English listeners actually overlook stress contrasts and treat such pairs as homophones (Cutler, 1986; Small et al., 1988).

But minimal pair distinction is not the only way in which information about lexical stress patterns can be useful in spoken-word recognition. Irrespective of language, recognizing spoken words is a continuous operation; incoming information in the speech signal is processed as it arrives and is used to favour matching interpretations of the signal and disfavour mismatching ones. This is efficient because of the structure of vocabularies, which all contain very large numbers of words – in the hundreds of thousands – constructed from only relatively few – on average 2–3 dozen – phonemes. Inevitably, then, words of any vocabulary will resemble one another (*insight*, *inside*), shorter words will occur accidentally within longer ones (*inn* and *sighed* in *inside*), and, especially with shorter words (*in*, *inn*; *cite*, *sight*, *site*; *side*, *sighed*), homophony will be the rule rather than the exception (Frost and Katz, 1992, for cross-language examples; Piantadosi et al., 2012, prove the computational efficacy of homophony). The structure of speech, equally inevitably, involves sequences with multiple interpretations (as in the young lady from Ryde who ate some green apples and died after they ‘made cider inside her inside’); the listener’s task is to sort out the words intended to form a message from the many alternative interpretations that are fully or partially supported by the acoustic signal.

In this processing task, information encoded suprasegmentally can be used in many ways in various languages. For instance, it can play a role in distinguishing segments; thus, final voicing distinctions in English (as in *incite* versus *inside*) are largely cued by

[1] This type of variable-placement stress is referred to in the literature as ‘free stress’ or ‘lexical stress’; we use the latter term.

[2] We follow Ladefoged and Johnson (2014, table 4.2) in using ‘full’ for vowels with full vowel quality, whether stressed or not, and ‘reduced’ for vowels that surface as [ə, ɨ] or similar central vowels.

[3] Upper case will henceforth represent primary stress location.

DOI: 10.1159/000447428

the duration of the preceding vowel (Raphael, 1972; Broersma, 2005). And certainly recognition of the difference between primary and secondary stress on a syllable can help distinguish words. Suppose some background noise masks the end of a friend's suggestion: *Let's meet outside the mu-*; if the suggestion is *outside the music school*, the last intelligible syllable would bear primary stress, whereas in *outside the museum* the same syllable would bear secondary stress. In this situation the stress information on the word's first syllable alone could sort out two likely continuations, and even in the absence of noise interference, these two words could be distinguished earlier than by waiting for the segment at which they diverge. So stress information could speed up the recognition of spoken words.

Nonetheless, there is abundant evidence that English-speaking listeners essentially ignore this type of information. Just as minimal stress pairs are treated as effectively being homophones (Cutler, 1986; Small et al., 1988), so is there no adverse effect on the speed of recognition of any word if stress is shifted (nota bene without altering segments) to create a non-existent word form – *stampede* pronounced as *STAMpede* (Slowiaczek, 1990), *nutmeg* pronounced *nutMEG* (Cutler and Clifton, 1984). Listeners presented with sentences in which a final word was presented only as a stress pattern (for instance, *The zookeeper fed the DAdada*) found a mismatching continuation such as *gorilla* (correct pronunciation *goRILla*) just as acceptable as a matching continuation (e.g. *elephant*; *ELephant*; Slowiaczek, 1991). This is a phenomenon peculiar to English; in the very closely related language Dutch, where both the phonological stress rules and the acoustic realisation of stress are virtually identical to those in English (Sluijter and van Heuven, 1996; Trommelen and Zonneveld, 1999), misstressing indeed inhibits spoken-word recognition (Cutler and Koster, 2000; van Leyden and van Heuven, 1996), and minimal stress pairs are indeed distinguished by listeners on the basis of suprasegmental cues (Cutler and Donselaar, 2001; Jongenburger and van Heuven, 1995).

Techniques for studying lexical access in progress confirm the difference between these two very similar languages. These techniques include cross-modal priming studies in which listeners heard neutral sentences ending in a truncated word (e.g. *the password for this week was ADmi-*) and made lexical decisions about items that were presented visually as the word-initial fragment occurred (with the crucial comparisons being made between words that matched the fragment in stress, e.g. *ADmiral*, or did not match it, e.g. *admiRAtion*, compared with a neutral control, e.g. *proposal*). In such studies, a match between word fragment and visually presented word should always facilitate recognition; the crucial question involves inhibition in the mismatch condition. If the suprasegmental cues to word identity are being exploited, then *ADmi-* will match only *admiral*, and recognition of *admiration* will be inhibited. Exactly this result was observed in Dutch (Van Donselaar et al., 2005), but in the English version of the experiment only match facilitation appeared; the inhibition was not observed (Cooper et al., 2002). Unlike Dutch listeners, English listeners did not exploit the suprasegmental cues that contrast the initial syllables of, for instance, *admiral* versus *admiration*. Dutch listeners can even outdo English listeners at distinguishing such English pairs (Cooper et al., 2002; Cutler et al., 2007; Cutler, 2009).

The reason for this behavioural difference lies in the relative pay-off offered by Dutch versus English stress cues. In English, most unstressed syllables actually have a reduced vowel. For English listeners, the vowel inventory effectively contains 'two kinds of vowels' (Bolinger, 1981): full vowels, where vowel quality is informative, and reduced vowels, differing minimally in vowel quality and far less informative.

Vowel reduction is however far less widespread in Dutch. Thus, the Dutch cross-modal priming experiment contained the pair *octopus-oktober* 'octopus-October', in which the vowels in the first two syllables are the same in both words. But in English they are not. The second syllable of English *octopus* contains the reduced vowel schwa, not the full vowel of the second syllable of *October*. Lexicostatistical examinations of the two vocabularies (van Heuven and Hagman, 1988; Cutler and Pasveer, 2006) show that pairs of segmentally identical word-initial fragments differing in stress are more common in Dutch than in English, and that in consequence, Dutch offers a significantly greater reduction of potential alternative interpretations of incoming speech signals when suprasegmental information is exploited. In making use of this type of information in Dutch but not in English, therefore, listeners are, as ever, behaving rationally in accord with the statistics of their language.

What further consequences might such a cross-language processing difference have? It is clear that the suprasegmental dimension of stress realisation is not attenuated in English; acoustic differences between syllables differing in the level of stress they bear are robust (see e.g. Cutler, 2005), and as just noted, Dutch listeners can make use of these acoustic cues as efficiently in English as in their native language. Also, recognition of English speech requires listeners to use information in the same suprasegmental dimensions in which lexical stress is realised; durational differences cue final voicing and phrase boundary placement, for example, and intonational contours are exploited to locate salient information or to process syntax (e.g. question vs. statement). Finally, English listeners are not 'stress deaf' in the manner of French, Finnish or Hungarian listeners (Dupoux et al., 1997; Peperkamp et al., 2010); it is hard for these groups to decide whether a nonsense token *BOpelo* matches an earlier token of *bopeLO* or *BOpelo*, but this is easy for speakers of English.

It seems, therefore, that it is only in the process of accessing lexically stored forms that English listeners fail to make use of suprasegmental information. But again, we can reasonably ask where this access process begins. Word recognition, as delineated above, is a process of sorting out alternative possible continuations of a gradually arriving input, with each partially or fully identified phoneme constraining the decision process by favouring some candidate words and disfavouring others. Segment identification is thus in the service of word recognition and an inherent part of the process. So is there any difference in the role played by suprasegmental cues associated with stress in the segmental processing undertaken by English and by Dutch listeners, respectively?

An answer to this question can be found in the detailed information about segmental processing across time provided in two recent very large data sets, for Dutch and for English, respectively. Each data set arose from a study charting listeners' best guesses at identifying gated fragments of every diphone possible in the relevant language – so each study included every phoneme of the language in question, in every preceding or following context in which it could legally occur (either within a word or across adjacent words). Each diphone was divided into (usually) 6 fragments, and listeners judged one sixth, one third, half, two thirds, five sixths or all of each diphone, resulting in a view of the uptake of all possible phonemic information for the language, across time. The Dutch study (Smits et al., 2003), with 18 listeners and 2,294 diphones, yielded a data set of nearly half a million phoneme categorisations, and the English study (Warner et al., 2014), with 20 listeners and 2,288 diphones, produced just over half a million data points.

The segments that are affected by stress are, primarily, the vowels. In each of these studies, vowels that occur in stressed versus unstressed forms in the relevant language

DOI: 10.1159/000447428

were presented in both forms. Thus, each consonant-vowel or vowel-consonant diphone with such vowels came in 2 stress versions, while the vowel-vowel diphone sets ran to 4 versions (stressed-stressed, stressed-unstressed, unstressed-stressed, and unstressed-unstressed), an English example for one such sequence (/i/ followed by /e/) covering occurrences in, for example, *agree-eighty*, *agree-eighteen*, *angry-eighty*, *angry-eighteen*. The data sets thus provide a rich source of evidence on the uptake of information both for stressed and for unstressed vowels in each language.

The response measure in each data set is the proportion of correct identifications at each point across a presented token. We can thus assess the relative accuracy of vowel identification in stressed versus unstressed form in each language, and also the relative speed with which (across the three fragments, or ‘gates’, in which each individual vowel was presented) listeners reached a correct identification. Will Dutch listeners’ experience with the need to process suprasegmental cues to stress, and with the range of vowel types in unstressed syllables that has prompted it, pay off in better performance with each vowel version? Or will their sensitivity to stress cues lead to differing performance across stressed versus unstressed tokens? Will English listeners’ neglect of the suprasegmental information pay off in rendering the realisation (stressed or unstressed) irrelevant to their identification performance, as they concentrate on spectral cues alone? Or will they, even in this segmentally focused task where word recognition is not called for, be hampered by their everyday word recognition experience, in which there is no need for vowels to be classified in terms of levels of stress? In this report, we address these questions by comparing the effects of the stress dimension on vowel identification performance in each of these data sets.

Methods

Materials

Materials for the English and Dutch studies are described in full detail in Warner et al. (2014) and in Smits et al. (2003). Each experiment assessed all possible sound sequences that could occur either within a word or across word boundaries in the relevant language. ‘Possible diphones’ included all combinations of stressed and unstressed vowels, and of vowels, and of consonants. Thus, the English set included diphones such as /fp/ (as in ‘off-putting’) and /eɪ/ (as in ‘say in’) as well as more obvious sequences like /pi/ and /ip/ (both as in ‘peep’), and as noted above, there were two versions of consonant-vowel and vowel-consonant diphones, with the vowel stressed or unstressed, respectively, and vowel-vowel diphones included all four possible combinations of stress.

Diphones were recorded in a minimal context, e.g. /aˈCVkə/ (some CV diphones) or /ˈabVˈVkə/, /ˈabVVˈke/, etc. (VV diphones), so as to make all diphones pronounceable for the speakers (e.g. because /pt/ is not a possible utterance onset in either language) and also to prevent the second segments from being utterance final (which could lead to excessive final lengthening effects on the timing of perceptual cues). The contexts further allowed the stress of preceding and following vowels to be manipulated to make it easier for the speaker to pronounce diphones with a specified stress pattern. Methods for the two experiments were matched, with only a few adjustments to the materials due to phonological considerations (e.g. Dutch schwa has no stressed equivalent, while English schwa and /ʌ/ form an unstressed-stressed pair). All decisions about materials are explained in Smits et al. (2003), Warner et al. (2005) and Warner et al. (2014)[4].

[4] Because of the size of the project, there was a very small number of exceptions to the usual procedures for specific diphones that do not influence the overall results, and these are documented in Smits et al. (2003) and Warner et al. (2014).

The vowel inventory used for the English experiment included /i, ɪ, e, ɛ, æ, a, ʌ~ə, o, u, ʊ, ɝ~ɚ, ai, au, oi/[5], with /ʌ/ and /ə/ considered a stressed/unstressed pair in this variety of the language (older editions of Ladefoged and Johnson, 2014, list this as one possible analysis, and Zsiga, 2013, pp. 62–63, also implies this) and /ɝ~ɚ/ similarly used to transcribe the stressed and unstressed rhotic vowel. The original vowel inventory for the Dutch study was /i, ɪ, e, ɛ, y, ʏ, œ, a, ɑ, o, ɔ, u, ɛi, œy, au, ə/, but /ʏ, ə/, though expected to be phonemically distinct, were treated as identical by listeners in that study (Smits et al., 2003). This led to low accuracy for these vowels, and they were excluded from further analysis (and thus are also absent from the present paper). This situation differs from English /ʌ~ə/ in that /ʏ/ (spelled 'u') can occur both stressed (e.g. *KUSsen* 'cushion') and unstressed (*ZEEkust* 'seaside', *TECHnicus* 'technician').

Each diphone was gated at 6 time points, with gate end points at one third, two thirds, and the entire duration of the first segment, and one third, two thirds, and the entire duration of the second segment. Stimuli were only final gated; any preceding context and all portions of the diphone up to the gate end point were included. For example, the gate 5 stimulus for the unstressed diphone /pe/, recorded with a preceding /a/, allowed the listener to hear /ˈape/ up to two thirds of the duration of the /e/. Thus, recorded following context was never included in any stimulus token, but preceding context was always included, and gates 4–6 (which end during the second segment) always included the first segment. Each stimulus was followed by a square wave beep sound, with the speech amplitude ramped down and the amplitude of the square wave simultaneously ramped up over a period of 5 ms starting at the gate end point. This procedure avoids artefactual perceptual effects of cutting speech suddenly to silence.

Subjects

Eighteen native Dutch listeners and 20 native American English listeners completed the studies. The Dutch subjects were students at the Radboud University, Nijmegen, the Netherlands; the English subjects were students in the Honors programme of the University of Arizona. All were monolingual in their native language (Netherlands Dutch or American English) until they began learning foreign languages in school. Due to the many ongoing vowel shifts in various US dialects, the English listeners were screened to ensure that they came from the same general dialect region as the speaker of the stimuli (southwestern USA, including southern California).

Procedures

For each language, the total set of stimuli was randomised and presented to subjects in sessions of approximately 1 h each, with breaks during a session. Subjects sat in a sound-protected booth, at the Max Planck Institute in Nijmegen and at the University of Arizona, respectively. They heard stimuli over high-quality headphones and used a mouse to click on symbols on the computer screen to indicate their responses for the two sounds of the diphone. They heard each stimulus once and then clicked on the left half of the screen for the first sound and on the right half of the screen for the second sound, with the same response options (all phonemes of the language) shown on each screen half. Each stimulus was presented just once because of the number of stimuli. For diphones with preceding context, that context was displayed at the left edge of the screen, outside the response option area, and subjects were instructed that such displays were not part of the 2-sound sequence they were identifying.

Subjects received training in the response symbols. Dutch orthography is consistent enough that it was possible to use response symbols based largely on Dutch spelling, with only a few sounds requiring special symbols. Because English orthography is inconsistent and there is no unambiguous way to write some vowels, the English response symbols diverged more from typical spelling than the Dutch symbols did. English subjects therefore received more training on the response symbols. For example, for English 'oo' was used as the response symbol for /u/ and 'uu' for /ʊ/. The task overall can be regarded as more difficult for English because of this, but accuracy of the English subjects (Honors students) was not substantially less than that of the Dutch subjects in most categories. Further details of response categories and how results were evaluated are available in Warner et al. (2014).

[5] The transcription system used for English vowels, especially diphthongs, has been modified slightly from the one used in Warner et al. (2014) in order to make it more similar to the transcription system used for Dutch. For example, in this paper we use /e/ instead of /ej/ and /au/ instead of /aw/.

Both groups first received a practice session using stimuli from the actual experiment, including strong representation of phonemes with response symbols that diverge from typical orthography. As described in the earlier papers, subjects were excluded if they performed badly on the practice session. Because of the unusually large number of stimuli, many of which sound similar especially for early gates, it was deemed to be unlikely that having heard a few hundred of the stimuli during the practice would influence the subjects' responses to them when hearing them again months later randomised among thousands of other stimuli.

Subjects in both studies returned to the lab for multiple sessions of about 1 h until they completed all stimuli. Dutch subjects had an average of 27.9 visits, English subjects an average of 32.73 visits. The difference is due to the greater difficulty of the English response symbols and scheduling around class times that made most visits for English subjects a little shorter than 1 h. However, all procedures were matched as closely as possible across studies.

Results

Following the earlier papers, we analyse for each language the proportion of correct identifications per gate for all stimuli containing a given vowel or vowel type (short, long, diphthong), in either the first (segment 1) or second position (segment 2) of the diphone. Thus, the data presented here is averaged over all the diphones containing a given vowel (e.g. stressed /i/) or vowel type (e.g. unstressed diphthongs) in each position. For example, for proportion correct for stressed /i/ as segment 1 at gate 3, responses from gate 3 of all diphones with stressed /i/ as the first segment are averaged, regardless of the nature of segment 2. All statistics reported are within-subject ANOVAs, with proportion of responses correct for a given segment converted to rationalised arcsine-transformed units (RAU; Sherbecoe and Studebaker, 2004) as the dependent variable, as recommended for proportion data.[6]

Comparison across Languages, by Vowel Type

Because the vowel inventories of Dutch and English differ, we cannot directly compare the effect of stress for each vowel. Some similar phonemes (e.g. /i/) exist in both languages, but there is neither an English match to Dutch /œ/ nor a Dutch match to English /æ/. However, both languages have a similar structure in the vowel space in that both have short or lax vowels, long or tense vowels, and diphthongs. English and Dutch also both have long/short pairs such as /i, ɪ/ and /e, ɛ/, with phonological analyses of each language dividing the vowels in this way. Since the English tense/lax distinction can also be described as long/short, it is reasonable to treat the vowel types short, long, and diphthong as comparable across the two languages. For Dutch, as for the analyses conducted by Warner et al. (2005), /ɑ, ɛ, ɪ, ɔ/ were classed as short, /a, i, u, y, e, o, œ/ as long, and /ɛi, œy, au/ as diphthong. For English, /ɪ, ɛ, æ, ʊ, ʌ, ə/ were classified as short/lax, /i, e, a, o, u, ɝ~ɚ/ as long/tense, and /ai, au, oi/ as diphthong.

[6] Using ANOVAs on arcsine-transformed proportions also avoids some problems that a linear mixed effect (LME) approach might present. First, the analysis of a categorical factor with 6 levels, such as gate, would be problematic to interpret with LME if treatment coding is used, as in many uses of LME. Second, the preferred analysis would be generalised LME, with a categorical dependent variable of correct/incorrect for each presentation of a stimulus. However, there is considerable structure within the set of diphones in a given condition (e.g. the set of diphones with stressed /i/ as segment 1), because of patterns in which phonemes are possible as segment 2 when segment 1 is being analysed, and vice versa. This would lead to problems of non-independence among the items. Note that items (diphones in this case) is not a true random factor.

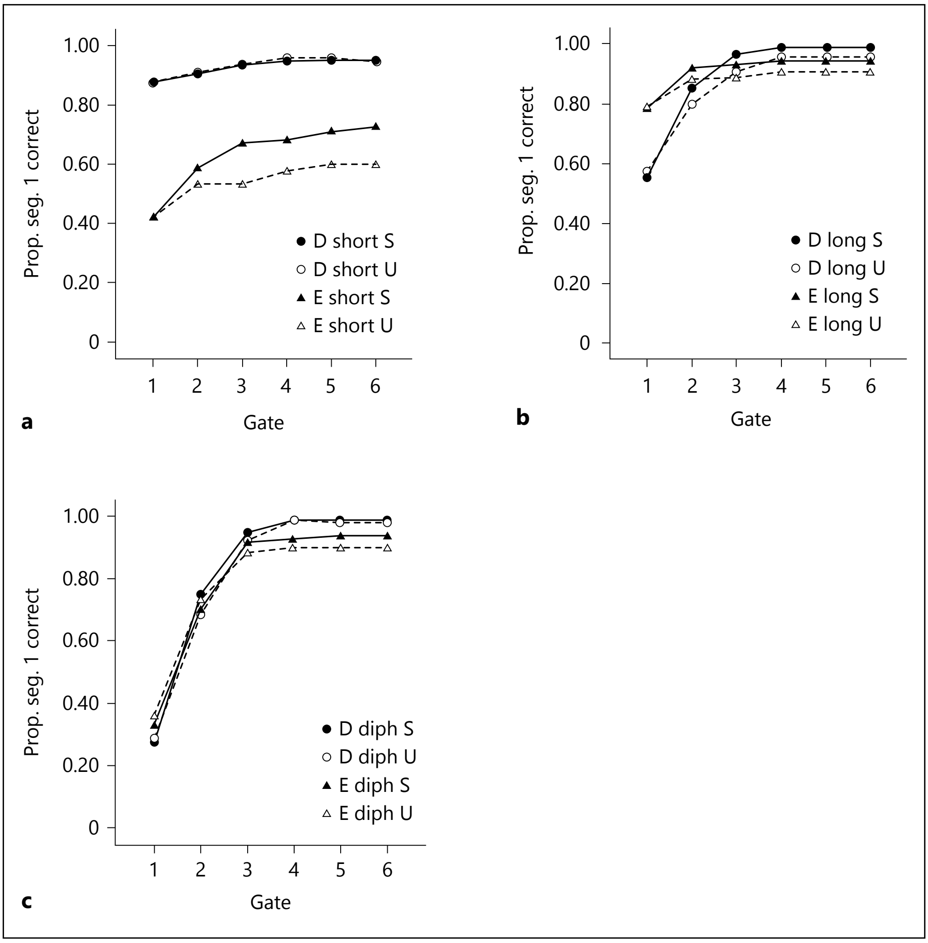

Fig. 1. Proportion of correct responses for segment 1 of all diphones with a vowel as segment 1, by language (D = Dutch, E = English), gate, and stress of the segment 1 vowel (S = stressed, U = unstressed). **a** Short/lax vowels. **b** Tense/long vowels. **c** Diphthongs.

Figures 1 (segment 1) and 2 (segment 2) display the overall average proportion correct across subjects by vowel type, language and stress across gates (i.e. thirds of each segment). The figures show that in many conditions, stressed vowels are perceived more accurately than unstressed ones, but that there are differences in how this effect of stress develops over time through the course of the diphone (i.e. across gates).

We analysed the data using a 4-factor ANOVA with the factors language (Dutch, English), gate (1–6), stress (stressed, unstressed), and vowel type (short, long, diphthong), with the proportion of segment 1 correct, transformed to RAU, as the dependent variable. (Segment 2 will be discussed below.) Language is a between-subjects factor, all other factors are within subject. For segment 1, all main effects and all interactions, including the 4-way interaction, were significant (all p values <0.001 unless noted):

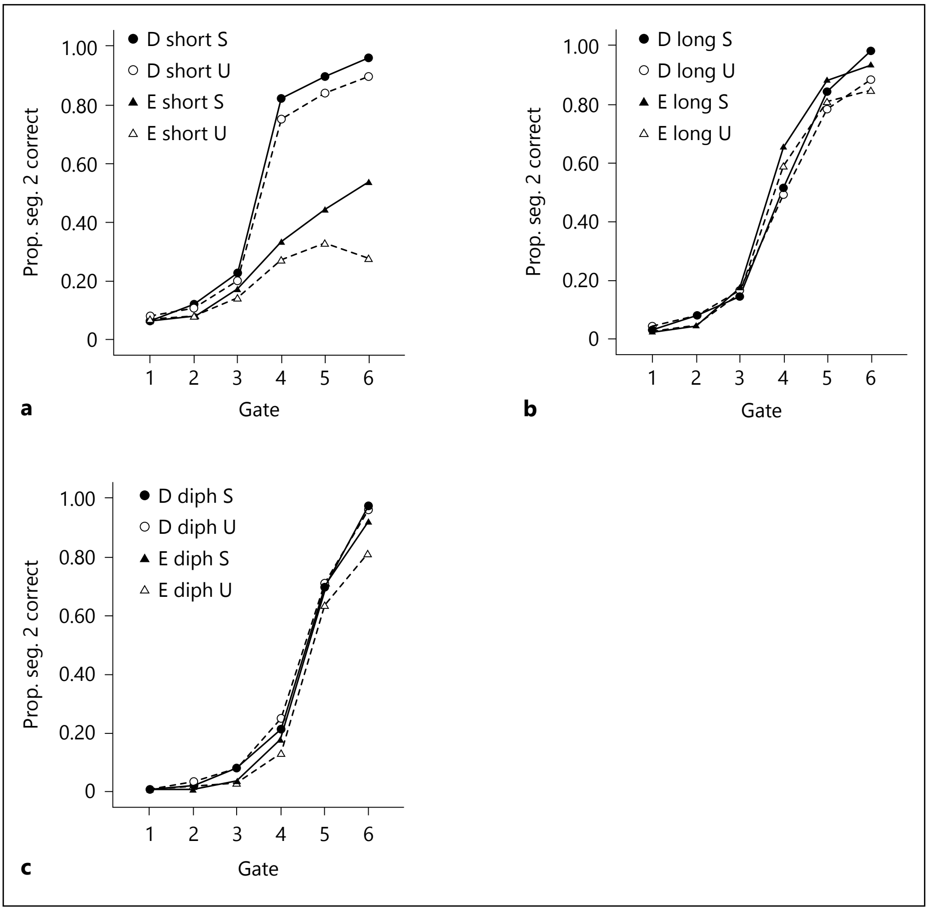

Fig. 2. Proportion of correct responses for segment 2 of all diphones with a vowel as segment 2, by language (D = Dutch, E = English), gate, and stress of the segment 2 vowel (S = stressed, U = unstressed). **a** Short/lax vowels. **b** Tense/long vowels. **c** Diphthongs.

language: F(1, 36) = 74.98; gate: F(5, 180) = 548.18; stress: F(1, 36) = 357.71; vowel type: F(2, 72) = 65.40; gate × language: F(5, 180) = 19.23; stress × language: F(1, 36) = 35.27; vowel type × language: F(2, 72) = 65.40; gate × stress: F(5, 180) = 41.23; gate × vowel type: F(10, 360) = 152.97; stress × vowel type: F(2, 72) = 16.44; gate × stress × language: F(5, 180) = 12.20; gate × vowel type × language: F(10, 360) = 24.10; stress × vowel type × language: F(2, 72) = 40.93; gate × stress × vowel type: F(10, 360) = 2.81, $p < 0.005$; 4-way interaction: F(10, 360) = 8.23. Thus, stress effects develop differently over time depending on the vowel type and language. Even though the higher-order interactions are significant, the 2-way interactions of stress by other factors are also informative: English overall shows a significantly larger stress effect than Dutch (language × stress), and the stress effect grows over time (gate × stress interaction).

To investigate these interactions, we tested the 3-factor comparison language × gate × stress for each vowel type separately. The 3-way interaction of these factors

Table 1. F ratios and significance of language (L), stress (S), and language × stress interaction (L × S) for each gate and vowel type, for segment 1

Gate	Short vowels	Long vowels	Diphthongs
1	L: 157.27*** S: F <1 L × S: F <1	L: 48.68*** S: 5.04* L × S: F <1	L: 1.06 S: 9.81** L × S: 3.97
2	L: 135.68*** S: 3.43 L × S: 5.23* E > D	L: 8.89** S: 114.64*** L × S: 2.59	L: F <1 S: 1.50 L × S: 27.45*** D > E
3	L: 179.67*** S: 56.32*** L × S: 52.89*** E > D	L: 8.38** S: 247.74*** L × S: 21.67*** D > E	L: 3.67 S: 39.35*** L × S: F <1
4	L: 206.23*** S: 24.03*** L × S: 76.49*** E > D	L: 60.78*** S: 135.82*** L × S: 6.04* E > D	L: 41.85*** S: 8.21** L × S: 5.47* E > D
5	L: 170.75*** S: 39.72*** L × S: 64.39*** E > D	L: 53.56*** S: 156.17*** L × S: F <1	L: 36.40*** S: 30.97*** L × S: 7.31* E > D
6	L: 141.11*** S: 48.47*** L × S: 46.28*** E > D	L: 51.89*** S: 153.43*** L × S: 1.27	L: 35.08*** S: 19.47*** L × S: 9.87** E > D

Degrees of freedom are 1 and 36 for each main effect and the interaction. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Direction of effect (which language shows the larger benefit for stressed vowels) is shown as E > D or D > E for each condition with a significant interaction.

was significant for each vowel type [short: F(5, 180) = 8.74, $p < 0.001$; long: F(5, 180) = 3.39, $p < 0.01$; diphthong: F(5, 180) = 14.26, $p < 0.001$]. Therefore, we tested the 2-way comparison language × stress at each gate for each vowel type (table 1). These 2-way comparisons address the central question we wish to answer: how does the size of the stress effect differ in Dutch versus English? In most cases where there is an interaction of language by stress, the stress effect is larger for English than for Dutch.

The results for the short vowels show no stress effect in either language at the first gate, when listeners have only heard the first third of segment 1. Thereafter, English develops a stress effect by gate 2 and maintains it in gates 3–6, while Dutch does not. Even by gate 6, when listeners have heard a full further segment after the vowel, English listeners' accuracy for unstressed vowels has not caught up to that of the stressed vowels.

For the long vowels, the two languages have a stress effect of similar size at most gates. They do differ significantly at gates 3 and 4 (the end of the vowel itself, and one third of the way through the following segment), but with opposite directions of the effect. Although significant, these interactions are quite small and only hold for 1 time point each. For example, at gate 4, Dutch has a difference in proportion correct for stressed versus unstressed of 0.033, while the difference for English is 0.037. (The significance of such a small interaction reflects the use of RAU as the dependent variable.

RAU is intended to stretch the 0.0–1.0 scale near its ends, and tends to make differences near ceiling – here, for Dutch – or near floor somewhat larger.) Apart from the small differences at gates 3 and 4, the two languages are comparable in the size of the stress effect for long vowels, with the stressed vowels perceived better by the second gate, and maintaining that advantage throughout the rest of the diphone.

For the diphthongs, gates 4–6 show a benefit for stressed vowels for English, with effectively no stress difference for Dutch. Thus, the stress effect develops later in English diphthongs than in short or long vowels. The significant interaction at gate 2 reflects a brief reversal in the direction of the stress effect for English, which probably stems from slight differences in the timing of diphthong transitions.

In summary, English shows a larger perceptual benefit than Dutch for stress on vowels (or a larger deleterious effect of lack of stress) for most of the duration of the short vowels and for all gates after the segment itself for the diphthongs. The few conditions with a larger stress effect for Dutch do not form a consistent pattern.

There are several other patterns in the results. The perception of English short vowels, especially unstressed ones, is quite poor. Even by the end of the following segment (gate 6), accuracy for English unstressed short vowels has not reached 60% correct, and even the stressed short vowels have low accuracy. (This poor performance does not appear to be caused by the more difficult English orthography, since the incorrect responses are not the ones that would be motivated by spelling-based misunderstandings. For example, /ʊ/ is most often misperceived as acoustically nearby /ə/, not as orthographically similar /u/ ‘oo’.) Dutch does not have this poor performance for short vowels, as shown by the significant interaction of language × gate × stress for short vowels. Recall, though, that the most poorly perceived Dutch short vowel, /ʏ/, was excluded since listeners could not distinguish it from schwa. The poor performance on English short vowels is not however confined to particular vowels.

Segment 2 vowels were analysed in the same way as those of segment 1. Again all main effects and interactions of the 4-factor ANOVA were significant: language: F(1, 36) = 128.83; gate: F(5, 180) = 4,251.61; stress: F(1, 36) = 328.83; vowel type: F(2, 72) = 57.14; gate × language: F(5, 180) = 61.06; stress × language: F(1, 36) = 59.36; vowel type × language: F(2, 72) = 53.28; gate × stress: F(5, 180) = 228.51; gate × vowel type: F(10, 360) = 194.76; stress × vowel type: F(2, 72) = 35.41; gate × stress × language: F(5, 180) = 6.75; gate × vowel type × language: F(10, 360) = 85.35; stress × vowel type × language: F(2, 72) = 10.79; gate × stress × vowel type: F(10, 360) = 10.78; 4-way interaction: F(10, 360) = 12.90. Tests of the language × gate × stress effect for each vowel type revealed significant 3-way interactions for each: short: F(5, 180) = 14.66; long: F(5, 180) = 8.75; diphthong: F(5, 180) = 8.01; all p values <0.001. Language × stress was thus again tested for each gate and vowel type (table 2).

As expected, identification of segment 2 is weak at early gates, where listeners have in fact only partially heard segment 1. With RAU, though, even small differences near the floor or ceiling are exaggerated and can show significance, as noted. This is the source of the significant language × stress interactions for short vowels at gate 1 and diphthongs at gates 1–2. At later gates, significant interactions usually reflect a larger advantage for stressed vowels in English than Dutch: for short vowels at gate 6, long vowels at gate 4, diphthongs at gates 4–6, and (very slightly) long vowels at gate 6. The only remaining significant interaction is for long vowels at gate 3, where the accuracy for all language × stress conditions is close, but Dutch shows a small reversal of the stress effect (not present at other gates), giving a significant interaction. The language ×

Table 2. F ratios and significance of language (L), stress (S), and language × stress interaction (L × S) for each gate and vowel type, for segment 2

Gate	Short vowels	Long vowels	Diphthongs
1	L: F <1 S: 9.20** L × S: 16.40***	L: 5.62* S: 1.38 L × S: 3.84	L: F <1 S: 2.74 L × S: 7.32*
2	L: 7.36* S: 2.01 L × S: F <1	L: 13.62** S: F <1 L × S: F <1	L: 47.39*** S: 14.96*** L × S: 5.45*
3	L: 10.14*** S: 31.44*** L × S: F <1	L: F <1 S: 1.52 L × S: 31.23***	L: 39.47*** S: 2.37 L × S: 3.32
4	L: 187.96*** S: 60.83*** L × S: 2.17	L: 22.30*** S: 69.93*** L × S: 22.54***	L: 7.15* S: 3.15 L × S: 50.90***
5	L: 159.19*** S: 114.69*** L × S: 2.21	L: 3.38 S: 117.36*** L × S: 2.04	L: 1.84 S: 10.67** L × S: 29.31***
6	L: 178.13*** S: 233.79*** L × S: 26.02***	L: 35.20*** S: 432.49*** L × S: 6.69*	L: 62.62*** S: 84.34*** L × S: 36.66***

Degrees of freedom are 1 and 36 for each main effect and the interaction. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. All significant interactions are in the direction of a larger advantage for stressed over unstressed vowels in English than in Dutch, with Dutch showing occasional small reversals (slightly greater accuracy for unstressed).

stress interactions overall show more stress effect in English than in Dutch for short vowels at their end (gate 6), for long vowels one third of the way through (after which Dutch and English stress effects are alike), and for diphthongs throughout the segment.

Thus, the perception of segment 1 and segment 2 of the diphone is similar. In both positions, the stress effect is especially strong for English short vowels. For long vowels, the size of the stress effect for the two languages is more similar at most time points, and although some interactions are significant, they do not represent a consistent pattern over time. For diphthongs, both segments show a larger stress effect for English than Dutch at several time points, but the timing of when this effect develops differs. This reflects the fact that segment 2 is always preceded by another segment, often a vowel, whereas most sounds in segment 1 position are utterance-initial or preceded by just /b/; little or no preceding context thus conveys co-articulatory cues to a vocalic segment 1.

One further pattern here is mentioned in Warner et al. (2014): the English short vowels and diphthongs in segment 2 position have a notable increase in the stress effect between gates 5 and 6. For the short vowels, this is so extreme that perception of segment 2 unstressed short vowels is actually less accurate at gate 6 than at gate 5, even though by gate 6 listeners have heard the whole vowel. The explanation for this lies in the following context that all diphones were recorded with, but which listeners did not hear; co-articulation with the upcoming consonant seems to be especially damaging to the perception of unstressed short vowels. The results suggest that this is less the case for Dutch, although scores for individual vowels (discussed below) differ.

The Stress Effect within Each Vowel of Each Language

Although grouping the vowels as short, long, and diphthong allows direct comparison across the two languages, it may obscure stress effects that are specific to one or a few vowels, rather than generalizing to an entire vowel type (short, long, diphthong). To investigate potential specific patterns, we tested stress × gate interactions for each vowel of each language separately. Because of the number of resulting comparisons (14 vowels in each language), we applied a Bonferroni correction (for 14 comparisons; critical p level = 0.00357), both to the 2-factor design and to tests for the simple effect of stress at each gate where there was a significant interaction. Results are displayed in figures 3 (segment 1) and 4 (segment 2), with statistical outcomes in tables 3 and 4.

The tables show a significant stress effect in more vowels in English than in Dutch (either the main effect in the stress × gate design or simple effects of stress at various gates), and the interactions discussed above showed that when both languages have a significant stress effect, it is typically larger in English. Comparison of figure 3a and b (short vowels) shows that although Dutch does have a significant stress effect for the short vowels /ɑ, ɔ/, it is much smaller than the effects for English short vowels. However, some English vowels also have no effect of stress, such as /ʊ/, which is poorly perceived whether stressed or not (Warner et al., 2014). One Dutch short vowel, /ɪ/, has an unexpected reversal of the stress effect, consistently from gate 3 on. Examination of confusion matrices based on the raw data reveals that Dutch listeners chose /i/ as a response more often for stressed /ɪ/ as segment 1 than for unstressed /ɪ/ as segment 1, causing the higher accuracy score for the unstressed vowel in this case.

In figure 3c and d (the long vowels), /a/ has the largest stress effect in both languages; several other long vowels show significant but less obvious stress effects. In figure 3e and f (the diphthongs; also see statistical results in table 3), English shows more widespread stress effects than Dutch. Dutch has a stress effect in only one diphthong, and then only at 1 gate (perhaps revealing a difference in timing of the diphthong transition rather than a stable stress effect). There is a significant stress effect in most gates of English diphones beginning /au/, from the end of the diphthong itself and then continuing throughout the following sound, and a brief effect for /ai/.

For vowels as segment 2, both languages have many significant stress effects, though effects are more widespread and larger for English than for Dutch. Both languages of course tend to develop any stress effect relatively late in the diphone, since segment 2 vowels only begin to be heard at gate 4. Significant stress effects before gate 4, especially at gate 3 (ending at the end of segment 1), could reflect stressed vowels spreading more co-articulatory cues into the preceding segment than unstressed ones do; even earlier small effects may be due to chance or to RAU-exaggerated differences near the floor.

Comparison of figure 4a and b confirms the pattern from segment 1 that English has larger and more widespread stress effects in short vowels than Dutch does, but that the effect of stress is vowel specific in both languages. In the long vowels, the

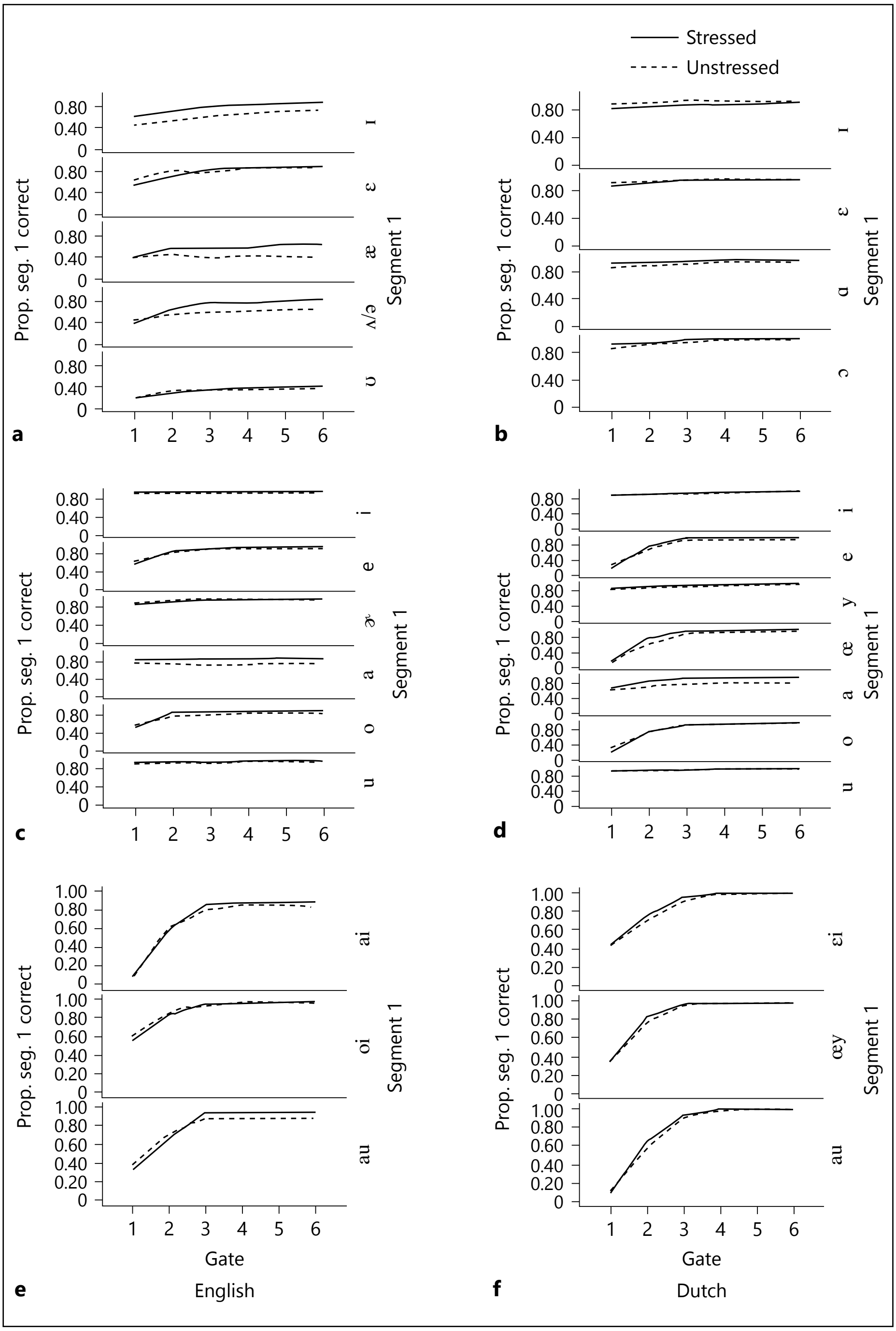

Fig. 3. Proportion of correct responses for segment 1 of all diphones with a vowel as segment 1, by vowel, gate, and stress of the segment 1 vowel. **a**, **b** English/Dutch short vowels. **c**, **d** English/Dutch long vowels. **e**, **f** English/Dutch diphthongs.

Table 3. F ratios for the 2-factor design of stress × gate for each vowel of each language as segment 1, and F ratios for the simple effect of stress at each gate if motivated by a significant interaction

Vowel	2-factor stress × gate			Simple effects of stress at each gate					
	stress	gate	inter.	1	2	3	4	5	6
English									
i	16.75	–	–						
ɪ	121.85	60.05	–						
e	–	52.73	–						
ɛ	–	44.19	8.50	–	–	–	–	–	–
æ	49.07	5.50	7.75	–	–	33.67	27.83	35.71	34.33
a	63.19	–	4.68	11.20	23.72	62.05	40.83	43.90	49.27
ʌ/ə	41.61	74.58	12.25	–	14.79	33.04	23.15	15.94	45.76
o	28.31	80.36	9.71	–	44.91	24.84	12.70	18.25	13.68
u	–	7.78	–						
ʊ	–	34.79	–						
ɝ/ɚ	–	42.63	5.85	–	12.95r	–	–	–	–
ai	–	317.75	4.37	–	–	–	–	–	25.67
au	33.48	137.53	21.04	–	–	30.25	39.73	72.62	28.67
oi	–	48.71	–						
Dutch									
i	–	8.00	–						
ɪ	33.12r	5.06	–						
e	27.19	211.00	18.53	12.88r	–	12.78	64.29	91.95	116.95
ɛ	–	–	–						
a	80.39	12.93	7.35	–	34.65	89.76	59.97	51.15	27.15
ɑ	15.85	4.39	–						
o	–	206.44	10.10	23.19r	–	–	–	–	–
ɔ	17.97	18.83	–						
u	–	15.84	–						
y	15.23	26.57	–						
œ	44.24	308.29	11.10	–	38.71	39.47	16.70	14.33	28.59
ɛi	–	55.27	4.23	–	–	24.91	–	–	–
œy	–	190.13	–						
au	–	279.48	–						

All effects are evaluated against a critical p level of 0.00357, based on the Bonferroni correction for tests of each of 14 vowels. For clarity, only significant F ratios are included, and a dash is used to denote an effect that was tested but was non-significant. Results marked with r have a reversal of direction of the effect, with unstressed perceived more accurately than stressed. All other significant effects of stress reflect more accurate perception of the stressed vowel. Degrees of freedom for English are 5 and 95 for the main effect of gate and the interaction, and 1 and 19 for all other tests. Degrees of freedom for Dutch are 5 and 85 for the main effect of gate and the interaction, and 1 and 17 for all other effects.

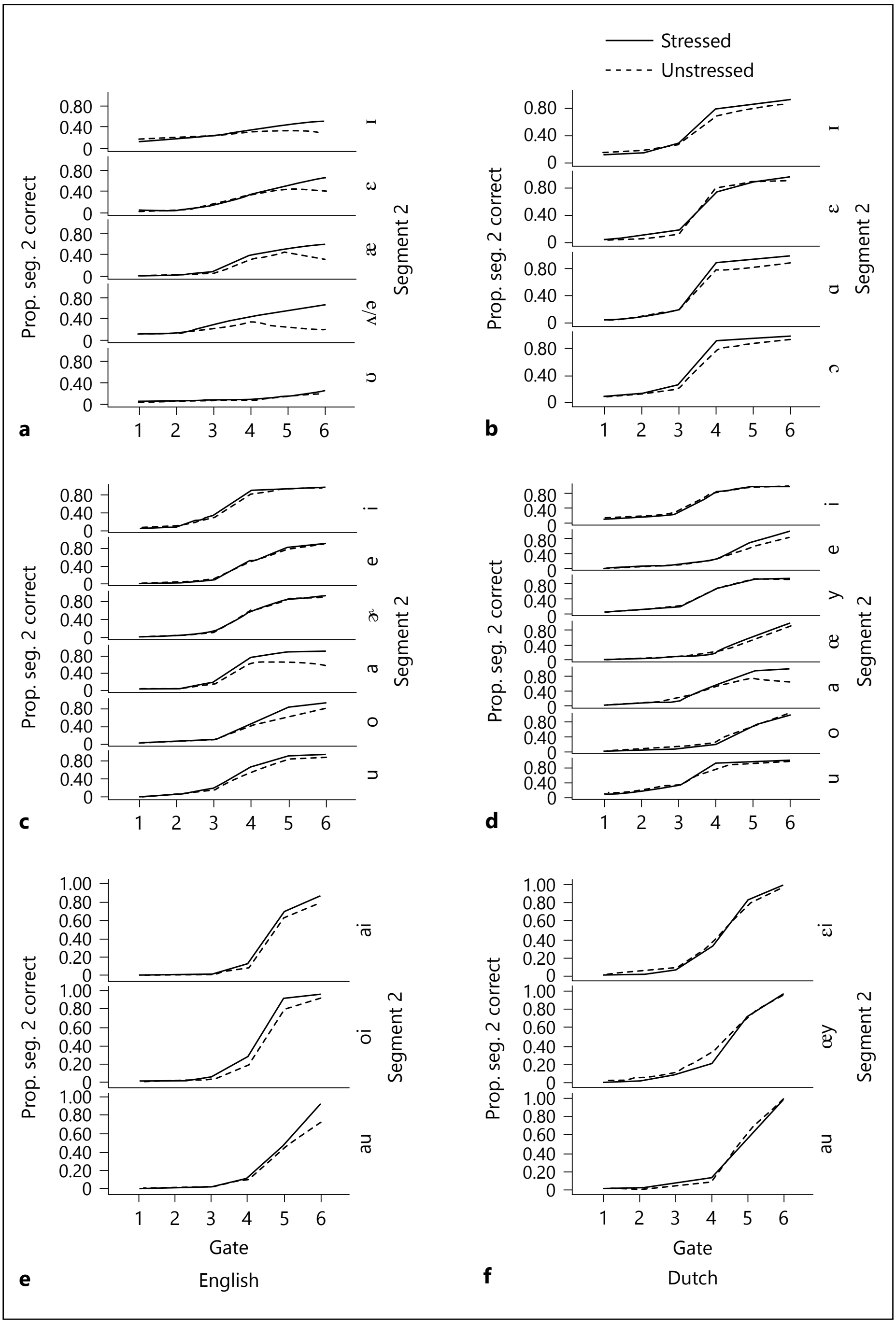

Fig. 4. Proportion of correct responses for segment 2 of all diphones with a vowel as segment 2, by vowel, gate, and stress of the segment 2 vowel. **a**, **b** English/Dutch short vowels. **c**, **d** English/Dutch long vowels. **e**, **f** English/Dutch diphthongs.

stress effect for both languages is again strongest for /a/, and the stress effect (if any) develops later for diphthongs than for other vowels. Warner et al. (2014) and Smits et al. (2003) both show that effects tend to develop one gate later for diphthongs than for other vowels, because of the delay before listeners have enough perceptual cues to perceive the second vowel quality of the diphthong. The English results for the short vowels as segment 2 (table 4, fig. 4a and b) again show a clearly increasing stress effect at the end of the diphone, from gate 5 to 6, again for all the short vowels except /ʊ/. For /ʌ~ə/, the increase in the stress effect begins even earlier (gate 4–5). In figure 4, this widening stress effect can be seen at gate 6 for the long vowel /a/ in both languages and for the diphthong /au/ in English, but it is not visible otherwise in Dutch. The widening stress effect at late gates reflects, as already noted, greater co-articulatory effects of following (but unheard) context on unstressed than on stressed vowels. This difference appears to affect English listeners more than Dutch listeners.

Comparison to Acoustic Measurements of Stimuli

Listeners of both languages perceive stressed vowels more accurately than unstressed ones, but this effect is stronger for English listeners than Dutch listeners. Although, as noted, evidence from prior studies points to similar acoustic realisation of syllabic stress in English and Dutch (Sluijter and van Heuven, 1996), there could still be a difference in how stress is realised in the materials of the present study (where different talkers produced the separate diphone sets). In particular, the English talker could have shifted unstressed vowels more, relative to stressed ones, while the Dutch talker's vowels changed little in quality with stress. This would yield less clear acoustic vowel quality information in the English stimuli. To test this, for the current study we measured the formant frequencies of the vowel stimuli of both earlier studies.

F1 and F2 were measured at each gate end point that fell during a vowel (hence at gates 1–3 for vowels as segment 1 and gates 4–6 for vowels as segment 2). For gates 1–2 and 4–5, the end point of the gate was used as the time point at which to measure the formants. For gates 3 and 6, formants were measured at 20 ms before the gate end point, because the gate end point itself is often exactly at the boundary between a vowel and a consonant, affecting the reliability of formant values measured at that point. Formant measurements were conducted automatically using a Praat (Boersma, 2001) script that used the 'To formant (Burg)' function, set to find 5 formants within the 0- to 5,500-Hz range. Examination of formant measurements as scatter plots of F1 × F2 verified that these automatic measurements were largely reliable. For both languages at most gate time points, a few tokens (approx. 2–5 out of over 1,000 stimuli) were outliers, with either F1 or F2 measurements outside the range of the rest of the data by more than 100 Hz. Such outliers usually arise when two formants are very close in frequency so that the automatic measurement mistakes one formant for another. These items were excluded from the data, or corrected if measurements had simply been shifted by one formant. Furthermore, the automatic measurement script failed to return formant values for a few stimuli (less than half of 1% of the total set), and these were excluded. Finally, for unknown reasons, the automatic measurement script returned implausible values (impossible within the human vowel space) for some Dutch diphones at gates 2 and 4, with no clear pattern as to which types of diphones are affected. Since none of the questions about the relationship of perception

DOI: 10.1159/000447428

Table 4. F ratios for the 2-factor design of stress × gate for each vowel of each language as segment 2, and F ratios for the simple effect of stress at each gate if motivated by a significant interaction

Vowel	2-factor stress × gate			Simple effects of stress at each gate					
	stress	gate	inter.	1	2	3	4	5	6
English									
i	33.34	976.59	7.74	–	–	–	39.81	–	–
ɪ	41.28	15.89	21.01	–	–	–	–	30.05	79.54
e	–	402.96	–						
ɛ	29.54	101.55	22.27	–	–	–	–	–	116.57
æ	72.89	100.16	28.07	–	–	–	15.98	–	72.85
a	269.59	484.97	48.63	–	–	16.78	24.97	100.38	321.01
ʌ/ə	51.38	51.03	61.19	–	–	–	–	61.45	124.85
o	54.19	505.34	30.74	–	–	–	–	87.55	76.51
u	66.89	1,017.31	4.40	–	–	11.65	39.75	14.81	13.31
ʊ	12.88	20.59	–						
ɝ/ɚ	–	1,385.74	4.75	–	–	–	–	–	–
ai	30.65	696.08	5.77	–	–	–	–	–	33.95
au	30.53	510.78	49.93	–	–	–	–	–	147.73
oi	139.19	852.41	11.55	–	–	–	38.82	79.83	17.62
Dutch									
i	–	1,415.45	3.93	–	–	–	–	–	–
ɪ	–	335.91	15.59	–	–	–	28.26	–	24.46
e	95.29	600.91	46.29	–	19.00	–	–	22.22	283.27
ɛ	11.43	869.37	14.26	–	43.79	24.40	–	–	18.71
a	60.66	416.75	49.61	–	–	–	–	42.87	86.31
ɑ	44.49	961.44	24.55	–	–	–	53.59	36.16	41.34
o	–	924.67	9.48	–	–	12.18r	12.29r	–	15.52
ɔ	40.73	1,100.53	21.89	–	–	17.63	123.30	26.06	26.77
u	–	1,463.29	13.77	–	–	–	84.32	–	–
y	–	1,193.10	–						
œ	62.35	727.84	18.29	–	–	–	–	29.62	43.51
ɛi	–	553.77	7.74	–	16.45r	–	–	–	–
œy	24.94	841.31	9.34	–	21.26r	–	73.44	–	–
au	–	1,146.46	10.28	–	–	38.69	–	22.05r	–

All effects are evaluated against a critical p level of 0.00357, based on the Bonferroni correction for tests of each of 14 vowels. For clarity, only significant F ratios are included, and a dash is used to denote an effect that was tested but was non-significant. Results marked with r have a reversal of direction of the effect, with unstressed perceived more accurately than stressed. All other significant effects of stressed reflect more accurate perception of the stressed vowel. Degrees of freedom for English are 5 and 95 for the main effect of gate and the interaction, and 1 and 19 for all other tests. Degrees of freedom for Dutch are 5 and 85 for the main effect of gate and the interaction, and 1 and 17 for all other effects.

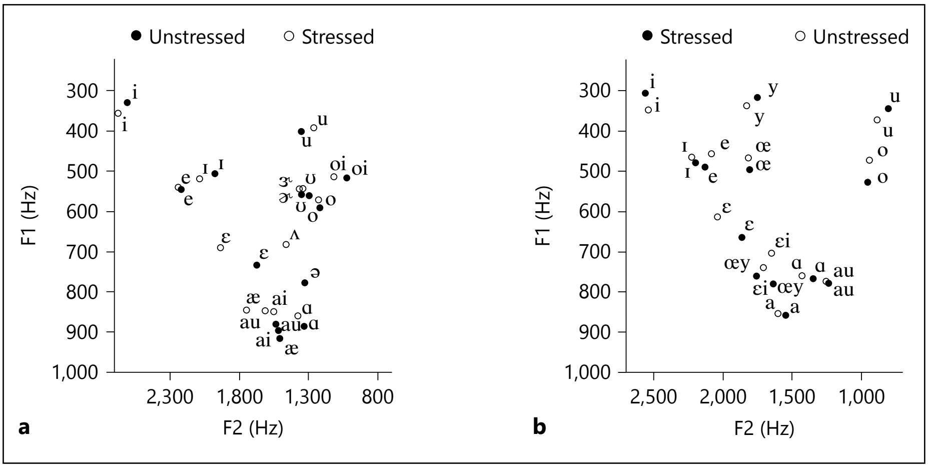

Fig. 5. Vowel space at the end point of gate 1, plotted as average F1 and F2 for each stressed and unstressed vowel. **a** English. **b** Dutch.

and acoustics hinge on these particular time points, we analysed the formant data only for the remaining 4 gates.

Figure 5 shows the average formant data for each vowel[7] analysed in the perceptual data above, for each language at gate 1 (at one third of the duration of the first segment). It can be seen from the vowel plots that both languages show some centralisation of unstressed vowels, especially in the F1 dimension. For English, this is especially evident in the vowels with higher F1 (low vowels and low-beginning diphthongs). For Dutch, a small degree of centralisation is also visible for high vowels. Both languages show little consistent effect of stress on location of the phonetically upper-mid vowels, and both languages show a large reversal of the expected centralisation effect for /ɛ/, with stressed /ɛ/ relatively central and low in the vowel space. (Since the pattern for this vowel for both languages is similar, it obviously cannot explain differences in the perceptual stress effect.) Vowel plots measured at other gate end points show similar patterns, with differences reflecting the transition during diphthongs (unstressed failing to move as far) or the transition to the following segment at gate 6.

We calculated the average size of the shift from stressed to unstressed for each vowel in each language at each gate (Euclidean distance in F1 by F2 space between the average for the stressed and the average for the unstressed vowel, averaged over all stimuli having that vowel as segment 1 or 2). Histograms revealed that both languages had vowel shift sizes in the same range except for the English vowel /ʌ~ə/, which had more shift at some gates than any Dutch vowel. We also calculated the average size of the perceptual advantage of stress for each vowel in each language at each gate, similarly by finding the difference between proportion correct (both raw proportion correct and RAU) when stressed and when unstressed (averaged over all stimuli having that

[7] Because of a coding error, formant measurements were not conducted for the Dutch vowel /ɔ/ in segment 1 position. This omission is unlikely to affect the results.

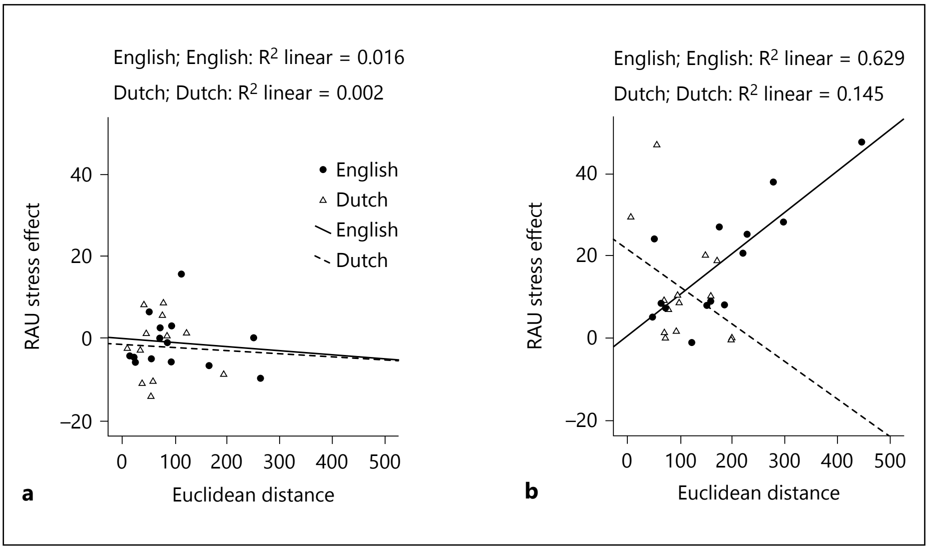

Fig. 6. Scatter plot of Euclidean distance between stressed and unstressed vowels versus amount of decrement in perceptual accuracy measured in RAU for unstressed vowels, with languages overlaid. **a** Gate 1. **b** Gate 6.

vowel as segment 1 or 2). We used Euclidean distance for the formant measurements, which is not directional, as a shift in any direction in the vowel space could be informative to listeners, and because an increase in F1 is centralisation for a high vowel, but not for a low vowel[8].

For each language at each gate, the pairwise correlation between the size of the perceptual stress effect and the size of the stress-conditioned shift in the vowel space was calculated. Results using RAU and raw proportion correct difference were largely similar, so we report the correlations with the RAU measure. Figure 6 shows scatter plots at 2 time points (gates 1 and 6)[9]. Correlations between the perceptual RAU difference and the acoustic Euclidean distance were non-significant for both languages at gate 1 (English: $r = –0.13$, $p > 0.05$; Dutch: $r = –0.05$, $p > 0.05$) and gate 3 (English: $r = 0.52$, $p > 0.05$; Dutch: $r = 0.09$, $p > 0.05$). At gate 5, the correlation was significant for English only (English: $r = 0.63$, $p < 0.02$; Dutch: $r = –0.03$, $p > 0.05$), but this English correlation was entirely due to the extreme value of the /ʌ~ə/ vowel. Without that vowel, the correlation for English was non-significant ($r = 0.26$, $p > 0.05$). However, at gate 6, English did show a significant correlation ($r = 0.79$, $p < 0.005$; without /ʌ~ə/: $r = 0.65$, $p < 0.02$), although Dutch did not ($r = –0.38$, $p >$

[8] This analysis looks for correlations between how large the acoustic difference and the perceptual difference between stressed and unstressed vowels are. It would also be possible to analyse the size of the vowel space for unstressed vowels across the two languages and look for correlations with average accuracy in perception of unstressed vowels, hypothesizing that a smaller vowel space is more confusable. Here, we focus instead on the size of the stress effect for each vowel.

[9] Because the speakers of the two languages use a similar range in both F1 and F2, and because correlations are calculated only within a language and hence within a single speaker, we did not apply any conversion for speaker normalisation but used raw hertz measurements instead.

0.05)[10]. The significant correlations were in the expected direction: greater acoustic difference between stressed and unstressed vowels correlated with greater perceptual decrement for unstressed vowels. The stronger correlations for English with the /ʌ~ə/ pair included reflect the fact that the stressed and unstressed vowel in this pair have a rather different quality. These vowels are not always considered to form a stressed/unstressed pair (Ladefoged and Johnson, 2014). Without that pair, the only significant correlation between perception and acoustics is for English at gate 6, at the end of the VC transition into the environment /k/ that followed all the vowel-final diphones in the recording materials.

Importantly, the English stimuli do not show a larger acoustic shift for unstressed vowels than the Dutch stimuli do. Just a few English vowel pairs, notably /ʌ~ə/, have a large shift. Furthermore, the Dutch listeners show no correlation between how much acoustic shift an unstressed vowel undergoes and how much perceptibility it loses. English listeners only reliably show such a relationship at gate 6, when all vowels are at the end of a VC transition into a following /k/ that the listener does not hear.

In summary, our comparisons of the perception results against the acoustic patterns in our stimuli have revealed that, in line with prior research, in most cases English unstressed vowels are not shifted to a greater extent than their Dutch counterparts. And at most time points (at all time points, for Dutch) we observed no systematic relation in the present data set between the acoustic and the perceptual effects of stress.

General Discussion

These two immense data sets (Smits et al., 2003; Warner et al., 2014), of around half a million data points each, have offered an unparalleled opportunity to view in detail the uptake of acoustic information for vowel identification. The present study exploited these resources to address a question motivated by the results of years of research on the role of stress in spoken-word recognition: listeners with English as a native language consistently fail to make good use of suprasegmental cues to lexical stress (e.g. to distinguish *music* from *museum* during the first syllable), while listeners with Dutch as native language efficiently exploit these acoustic cues.

Out of the richness of data our analyses provided, encompassing a view of every full vowel of each language with an existing unstressed or stressed realisation, placed in every possible context it could be encountered in, a clear message emerged: English listeners' vowel identification performance is subject to much greater effects of stress variation than is the performance of Dutch listeners. Stress effects – defined as better identification performance for a stressed as opposed to an unstressed realisation of the same vowel – were more widespread in our analyses and often occurred earlier and lasted longer, in English compared to Dutch. It is clearly not the case that English listeners' failure to attend to suprasegmental stress cues in making word recognition

[10] Correlation of the Euclidean distance with the difference in proportion correct rather than RAU had one additional significant correlation: English gate 3: $r = 0.54$, $p = 0.045$; however, without /ʌ~ə/ this correlation is not significant either: $r = 0.52$, $p > 0.05$. Otherwise, correlations of the acoustic measure with the proportion correct measure agreed in significance with the correlation with the RAU measure. Gates 2 and 4 are not reported here because of the problem with formant measurement for the Dutch vowels at those time points mentioned above.

decisions or metalinguistic decisions about speech allows them to factor out these cues and attend solely to spectral cues to vowel identity in a task requiring simple vowel categorisation. When the cues they want to attend to (spectral cues) are rendered more versus less clear, listeners are not immune to this variation; quite the reverse, they are more affected by this suprasegmental effect than Dutch listeners are.

As was clear from earlier work (particularly from studies in which Dutch listeners outperformed English listeners in identifying the stress level of English syllables such as *mu-* from *music* versus *museum*; Cooper et al., 2002) and was confirmed again in the analyses presented here, the acoustic differences between stressed and unstressed versions of the same vowel are quite similar in these two languages. It might therefore seem wantonly negligent of English listeners to ignore this informative dimension in interpreting speech signals. But as the research reviewed in the introduction abundantly shows, ignore it they do. Changing suprasegmental realisation alone does not impact upon word recognition success for English listeners; only changing segmental structure (by turning one full vowel into another, or altering a full vowel to a reduced one or vice versa) has such an impact (Bond and Small, 1983; Cutler and Clifton, 1984; Slowiaczek, 1990). Cross-splicing words such as to swap stressed and unstressed versions of the same full vowel has no effect on how natural a word sounds to English listeners (*autumn* with the first vowel of *audition*, or vice versa, are rated as being as acceptable as the original versions), though swapping a full vowel for a reduced one is judged unnatural (e.g. cross-splicing the first vowels of *audition* and *addition;* Fear et al., 1995). Likewise, 3-syllable rhythmic patterns are accepted as a match to 3-syllable words such as *elephant* and *gorilla* irrespective of where the stress is placed (Slowiaczek, 1991). Minimal stress pairs such as *insight-incite* are effectively treated as homophones (Cutler, 1986; Small et al., 1988), and stress-differing fragments such as the initial two syllables of *admiral* versus *admiration* fail to exercise inhibition on whichever word they mismatch (Cooper et al., 2002). The evidence is abundant and fully persuasive: wherever metalinguistic decisions are required, English listeners treat the English vowels not as a single set of vowels each of which may be realised as stressed or unstressed, but as one class of full vowels (all spectrally distinct from one another) and another class of reduced vowels (also spectrally distinguishable from one another, albeit rather less clearly).

Given such convincing evidence, it is perhaps no surprise that this decision on the part of English listeners is actually the most rational choice and entirely consistent with the prevailing view of listeners as Bayesian ideal observers (Norris et al., 2016). English listeners' prior experience has taught them that situations such as choosing between *music* and *museum* happen quite rarely, because most unstressed syllables in English are reduced to [ə] or a similar central vowel. This means that words tend to be less similar than in a language without vowel reduction or with less use of the vowel reduction option. Adjusted for word frequency, English polysyllabic words contain on average 0.9 embedded words if stress is not taken into account and 0.6 embedded words if stress is factored into the computation (Cutler et al., 2004). From the listener's point of view, that is on average one competing embedded word whether or not one takes stress into account. The sum is different in Dutch: 1.6 without taking stress into account, 0.8 if it is included in the computation, in other words a shift from more than one to on average just one (Cutler and Pasveer, 2006). As an ideal listener, then, the rational choice for the Dutch listener is to attend to suprasegmentals, while the English listener is better served by not doing so. Of course, we possess no direct estimate of the

DOI: 10.1159/000447428

actual added cost of attending to suprasegmental as well as to segmental information in recognizing spoken words; but this requires attending to separate acoustic dimensions (pitch, intensity) distinct from those used to distinguish vowel quality, and integrating that information with the vowel quality information. The evidence suggests it to be sufficiently substantial for the English listener not to waste it for such limited pay-offs.

In the light of English listeners' prior experience concerning vowels, the task in the diphone identification study of Warner et al. (2014) thus presented a lesser match to what might have been expected than did the Dutch study of Smits et al. (2003). Both studies presented every segment in every possible context as if each segment were equally likely, and equally likely in any context; this runs counter to expectation anyway, but in the case of vowels, more so for the English than for the Dutch listeners. A substantial proportion of the vowels that were presented were unstressed but not reduced. These vowels indeed occur in English (otherwise they would not have been in the stimulus set), but they do not occur in nearly as many words as stressed vowels or reduced unstressed vowels; moreover, the words they occur in may on average be lower in frequency than words with the more expected pattern (McQueen et al., 1995).

In contrast, Dutch listeners' performance showed much less stress effect, and this, too, can be explained in terms of their listening experience. Dutch is much less inclined to the liberal application of vowel reduction, so that Dutch vowels are realised far more often than English vowels in an unstressed but full-quality form. Considering cognate forms that are essentially the same in English and Dutch easily makes this point: *cobra* and *zebra* are stressed on the first syllable in both languages, but both second syllables have schwa in English, /a/ in Dutch. Likewise, *guitar*/*gitaar* and *cigar*/*sigaar* are all stressed on the second syllable, but where the first syllables have schwa in English, in Dutch they both have /i/. Thus, Dutch listeners' language has provided them with much more experience of full vowels in an unstressed realisation, which leads them to exhibit greater resilience to the variation in realisation across vowel tokens in this experiment. It also leads them to make full use of their experience in everyday speech recognition. In the same kind of experiments as show English listeners to ignore suprasegmental cues, Dutch listeners clearly attend to them; suprasegmental change slows Dutch word recognition (van Leyden and van Heuven, 1996), minimal stress pairs are treated as distinct (Jongenburger and van Heuven, 1995), and suprasegmentally distinct fragments such as *OCto-* versus *okTO-* selectively produce activation of only *octopus* or *oktober*, respectively (Van Donselaar et al., 2005). Dutch listeners even successfully apply this ability to making distinctions in English words that are ignored by English listeners (Cooper et al., 2002; Cutler, 2009).

Even without taking frequency of occurrence into account, the dictionary itself reveals a significant difference between these two (otherwise highly similar) languages in this respect. In the CELEX lexical database (Baayen et al., 1993), we tallied the relative occurrence frequency of all the vowels that can occur either with primary stress or without it, for the English and the Dutch dictionaries, respectively (in each case excluding the few vowels used rarely and only in loan words, and not included in the present analyses). Note that as each word can only have 1 vowel with primary stress, yet most words in each dictionary have 2 or more syllables, the prima facie expectation is for there to be a majority of vowel occurrences without primary stress. However, English manages to belie this: 51.15% of the tallied occurrences in the English lexicon had primary stress, and 48.85% had not, whereas in Dutch the primary-stressed vowels were indeed in the minority: 44.34% as against 55.66%

without primary stress. In both languages, but especially in English, certain vowels are particularly biased towards occurring in one or the other form. Nonetheless, even the dictionary resources of each language attest the lesser likelihood of unstressed full vowels in English.

Faced, then, with unstressed full vowels and required to identify them, the English listeners perform less well than they do with the stressed versions of the same vowels. In particular, they perform worse with the short vowels, where there is simply less acoustic evidence on which to base an identification decision. It is worth stressing that each of the original data sets presents a highly orderly outcome (see figure 1 in each of the papers by Smits et al., 2003, and Warner et al., 2014). Thus, the English listeners were unquestionably doing their best to identify the input. With the vowels, though, their degree of success in this was constrained by the relative familiarity of the realisation and by the amount of evidence available. Recall too that the English listeners still show a stress effect at gate 6 for vowels as segment 1 of the diphone. Even hearing potentially useful co-articulatory effects of the following context did not suffice to allow these listeners to compensate for the lesser information available in unstressed vowels.

As already noted, our failure to find any systematic relationship between the acoustic and the perceptual effect of stress and our confirmation of the similarity between English and Dutch with respect to the degree of acoustic shift of unstressed vowels combine to rule out any claim that a difference in the acoustic realisation of stress in the two languages might cause English listeners to experience a greater decrement in perception of unstressed vowels than Dutch listeners do. Instead, the explanation lies, as we have argued, in differences in the lexicon of the two languages.

We call attention here, however, to one as yet unexplained pattern in the English results, namely for the final fragment of vowel-second diphones (i.e. at gate 6). Recall that because of the following /k/ that was recorded to avoid sudden offset to silence, gate 6 offers a certain amount of co-articulatory information (in principle useful, though also perhaps misleading in the sense that it is not followed up with confirmation). The puzzling pattern revealed in the perception data is that the stress effect actually increases at gate 6, with perception of unstressed vowels sometimes becoming worse rather than better from gate 5 to 6. As figure 4 shows, such a pattern is displayed by about half of the vowels of English but by only one Dutch vowel. We suggest that this finding could conceivably reflect an acoustic difference between the languages, and therefore in consequence a perception/acoustics relationship, whereby co-articulation with the following context /k/ affected the unstressed vowels of English more strongly than those of Dutch, and hence impacted more on the English listeners' performance. The acoustic measurements give some suggestion of greater acoustic shift of English vowels than Dutch specifically for gate 6, which would support this. The correlation between perceptual and acoustic results for English for this particular gate (gate 6 for segment 2) seems to support this conclusion. The overall larger perceptual stress effect in English throughout the diphone is however clearly not an effect of greater vowel reduction, and must be sought in listeners' language-specific perceptual strategies rather than in the input acoustics.

In conclusion, it is clear that identification of vowels is affected by suprasegmental factors for any listener. English-speaking listeners cannot extract an advantage by ignoring this dimension in vowel processing, even though the word candidates that their vowel processing will lead them to consider in normal listening are not sorted by match or mismatch to the suprasegmental information received. Hearing unstressed *mu-*, they will process it slightly less efficiently than stressed *mu-*, but in both cases the vowel is full and not reduced, so either input will lead them to consider both *music* and *museum* as potential lexical interpretations. Only in the full versus reduced distinction do stress differences direct English listeners' processing; suprasegmental differences alone will be ignored. This reflects the fact that cases such as *music*/*museum*, where differing likely continuations do indeed present themselves, are quite rare. Dutch listeners, on the other hand, are used to being presented with alternative lexical continuations that differ in a particular syllable being stressed versus unstressed, so they must be able to interpret the suprasegmental stress cues; in Dutch, an unstressed *mu-* would produce *museum* as a potential word, and also, as it happens, *muziek* 'music', which in Dutch has final stress, but it would not produce *musicus* 'musician', in which the primary stress falls on the first syllable.

The English listeners' strategy of excluding the suprasegmental information from the spoken word recognition process, even though this information is present in the acoustics, may seem inefficient. The information is in the signal, so why not use it? However, like the Dutch listeners' strategy of using the same information in the same type of processing, this strategy is fully warranted by the probabilities associated with the relative recognition tasks for the two languages.

Acknowledgements

This research was supported by the Max Planck Society through employment of both authors at the time of the original Dutch study, support of the English study via a dedicated small grant, and provision of accommodation for the first author during the writing of the present paper. We are grateful to Sam Johnston for assistance with the formant analyses, to Priscilla Shin and Maureen Hoffman for their work on the English experiment, to Laurence Bruggeman for computing CELEX statistics, and to Benjamin Tucker for assistance with figures, and to James McQueen for useful discussions. We also thank the anonymous reviewers for their helpful comments. A brief report of part of the content of this paper was presented to the 166th meeting of the Acoustical Society of America, San Francisco, in December 2013, and we further thank Priscilla Shin and Maureen Hoffman for their contributions to that presentation.

References

Baayen RH, Piepenbrock R, van Rijn H (1993): The CELEX Lexical Database (CD-ROM). Philadelphia, Linguistic Data Consortium.

Boersma P (2001): Praat, a system for doing phonetics by computer. Glot Int 5:341–345.

Broersma M (2005): Perception of familiar contrasts in unfamiliar positions. J Acoust Soc Am 117:3890–3901.

Bolinger DL (1981): Two Kinds of Vowels, Two Kinds of Rhythm. Bloomington, Indiana University Linguistics Club.

Bond ZS, Small LH (1983): Voicing, vowel, and stress mispronunciations in continuous speech. Percept Psychophys 34:470–474.

Cooper N, Cutler A, Wales R (2002): Constraints of lexical stress on lexical access in English: evidence from native and nonnative listeners. Lang Speech 45:207–228.

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

Cutler A (2005): Lexical stress; in Pisoni DB, Remez RE (eds): The Handbook of Speech Perception. Oxford, Blackwell, pp 264–289.

Cutler A (2009): Greater sensitivity to prosodic goodness in non-native than in native listeners. J Acoust Soc Am 125:3522–3525.

Cutler A, Clifton C (1984): The use of prosodic information in word recognition; in Bouma H, Bouwhuis DG (eds): Attention and Performance X: Control of Language Processes. Hillsdale, Erlbaum, pp 183–196.

Cutler A, Koster M (2000): Stress and lexical activation in Dutch; in Yuan B, Huang T, Tang X (eds): Proceedings of the 6th International Conference on Spoken Language Processing. Beijing, China Military Friendship Publishing, vol 1, pp 593–596.

Cutler A, Norris D, Sebastián-Gallés N (2004): Phonemic repertoire and similarity within the vocabulary; in Kim S, Bae M (eds): Proceedings of the 8th International Conference on Spoken Language Processing, Jeju Island, Korea. Seoul, Sunjin Printing Co, vol 1, pp 65–68.

Cutler A, Pasveer D (2006): Explaining cross-linguistic differences in effects of lexical stress on spoken-word recognition; in Hoffmann R, Mixdorff H (eds): Proceedings of the 3rd International Conference on Speech Prosody. Dresden, TUDpress, pp 237–240.

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

Cutler A, Wales R, Cooper N, Janssen J (2007): Dutch listeners' use of suprasegmental cues to English stress; in Trouvain J, Barry WJ (eds): Proceedings of the 16th International Congress of Phonetic Sciences. Saarbrücken, Pirrot, pp 1913–1916.

Dupoux E, Pallier C, Sebastián-Gallés N, Mehler J (1997): A destressing 'deafness' in French? J Mem Lang 36:406–421.

Fear BD, Cutler A, Butterfield S (1995): The strong/weak syllable distinction in English. J Acoust Soc Am 97:1893–1904.

Frost R, Katz L (1992): The reading process is different for different orthographies: the orthographic depth hypothesis. Haskins Lab Status Rep Speech Res SR111/112:140–160.

Jongenburger W, van Heuven VJ (1995): The role of lexical stress in the recognition of spoken words: prelexical of postlexical? Proc 13th Int Congr Phonet Sci, Stockholm, vol 4, pp 368–371.

Ladefoged P, Johnson K (2014): A Course in Phonetics, ed 7. Nelson Education. Stamford, Cengage Learning.

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

McQueen JM, Cutler A, Briscoe T, Norris D (1995): Models of continuous speech recognition and the contents of the vocabulary. Lang Cogn Process 10:309–331.

Norris D, McQueen JM, Cutler A (2016): Prediction, Bayesian inference and feedback in speech recognition. Lang Cogn Neurosci 31:4–18.

Peperkamp S, Vendelin I, Dupoux E (2010): Perception of predictable stress: a cross-linguistic investigation. J Phon 38:422–430.

Piantadosi ST, Tily H, Gibson E (2012): The communicative function of ambiguity in language. Cognition 122:280–291.

Raphael LJ (1972): Preceding vowel duration as a cue to the perception of the voicing characteristic of word-final consonants in American English. J Acoust Soc Am 51:1296–1303.

Sherbecoe RL, Studebaker GA (2004): Supplementary formulas and tables for calculating and interconverting speech recognition scores in transformed arcsine units. Int J Audiol 43:442–448.

Slowiaczek LM (1990): Effects of lexical stress in auditory word recognition. Lang Speech 33:47–68.

Slowiaczek LM (1991): Stress and context in auditory word recognition. J Psycholinguist Res 20:465–481.

Sluijter AMC, van Heuven VJ (1996): Acoustic correlates of linguistic stress and accent in Dutch and American English. Proc 4th Int Congr Spoken Lang Process, Philadelphia, pp 630–633.

Small LH, Simon SD, Goldberg JS (1988): Lexical stress and lexical access: homographs versus nonhomographs. Percept Psychophys 44:272–280.

Smits R, Warner N, McQueen JM, Cutler A (2003): Unfolding of phonetic information over time: a database of Dutch diphone perception. J Acoust Soc Am 113:563–574.

Trommelen M, Zonneveld W (1999): Word-stress in West-Germanic: English and Dutch; in van der Hulst H (ed): Word Prosodic Systems in the Languages of Europe. Berlin, Mouton de Gruyter, pp 478–515.

Van Donselaar W, Koster M, Cutler A (2005): Exploring the role of lexical stress in lexical recognition. Q J Exp Psychol 58A:251–273.

Van Heuven VJ, Hagman P (1988): Lexical statistics and spoken word recognition in Dutch; in Coopmans P, Hulk A (eds): Linguistics in the Netherlands 1988. Dordrecht, Foris, pp 59–68.

Van Leyden K, van Heuven VJ (1996): Lexical stress and spoken word recognition: Dutch vs English; in Cremers C, den Dikken M (eds): Linguistics in the Netherlands 1996. Amsterdam, Benjamins, pp 159–170.

Warner N, McQueen JM, Cutler A (2014): Tracking perception of the sounds of English. J Acoust Soc Am 135:2295–3006.

Warner N, Smits R, McQueen JM, Cutler A (2005): Phonological and statistical effects on timing of speech perception: insights from a database of Dutch diphone perception. Speech Commun 46:53–72.

Zsiga EC (2013): The Sounds of Language: An Introduction to Phonetics and Phonology. Malden, Wiley-Blackwell.