

Dutch and English listeners' interpretation of vowel duration

Suzanne V. H. van der Feest^{a)}

Department of Communication Sciences and Disorders, The University of Texas at Austin,
1 University Station, Austin, Texas 78712
suzanne@austin.utexas.edu

Daniel Swingley

Department of Psychology, University of Pennsylvania, 3401 Walnut Street,
Philadelphia, Pennsylvania 19104
swingley@psych.upenn.edu

Abstract: Dutch and English listeners' interpretation of vowel duration changes was examined in a word transcription task. Listeners were presented with spoken words realized with canonical or altered vowel durations. Dutch listeners often misperceived lengthened short vowels and shortened long vowels, identifying them as the short/long counterpart of the target, whereas English listeners more rarely misidentified words with altered vowel duration. Although Dutch and English are similar prosodically and phonologically, listeners' treatment of vowel duration in clear speech is different across the two languages.

© 2011 Acoustical Society of America

PACS numbers: 43.71.Hw, 43.71.Es, 43.71.Sy [JH]

Date Received: September 3, 2010 Date Accepted: November 23, 2010

1. Introduction

Vowel duration depends on speaking rate, stress, and utterance position, among other factors (Klatt, 1976). In some languages, such as Japanese, vowel duration is described as *contrastive*: Several pairs of vowels are distinguished from one another almost entirely by duration. In other languages, such as Spanish, phonological vowel distinctions are not signaled by duration. Dutch and English, the languages we examine here, are less clear cases. In phonological descriptions Dutch is considered to maintain a featural vowel duration contrast, principally because of the rules that depend upon it: For example, the diminutive suffix takes a different form after a short (-*etje*) than after a long-vowel syllable (-*tje*; Moulton, 1962; Booij, 1995). Also, Dutch orthography, in contrast to English, clearly marks durational oppositions, as in *maan* ("moon") and *man* ("man"), rules that are very familiar to literate Dutch adults. Finally, Dutch unlike English speakers have been reported to be reluctant to exaggerate the duration of short vowels to convey emphasis in child-directed speech (Dietrich *et al.*, 2007). In most dialects of English vowel duration is not considered contrastive, though vowels vary in their "intrinsic" durations (House, 1961; Chomsky and Halle, 1968; Hillenbrand *et al.*, 2000) and duration varies depending on context, as will be discussed in more detail below. However, from a phonetic viewpoint, English and Dutch are alike in their opposition of pairs of long (tense) and short (lax) vowels, in which members of a pair differ (prototypically) in both quantity and quality. As a result, it is not clear whether the differential phonological descriptions of Dutch as maintaining a durational contrast and English as not doing so have a basis in native listeners' cue weighting. Here, we examined native-speaker identification of English and Dutch monosyllabic words varying in their vowel duration. Our goal was to determine whether Dutch listeners would use vowel duration to a greater degree than English listeners.

In English, coda voicing may affect vowel duration more strongly than the tense/lax distinction does (Denes, 1955). Hillenbrand *et al.* (1995) reported an average

^{a)}Author to whom correspondence should be addressed.

1.4 × ratio between tense and lax English vowels in isolated /hVd/ syllables; Raphael (1972) reported a 1.8 × ratio between voiced- and unvoiced-coda syllables for singleton codas. In Dutch, the voicing contrast is neutralized in word final position (Booij, 1995). Although Dutch speakers do not risk signaling the “wrong” coda voicing by realizing a vowel with too long or short a duration, the Dutch phonetic implementation of duration is described as more rigid than in English: Especially in utterance-final stressed syllables (the context we are investigating here) long vowels are about twice as long as short vowels. Dutch long-short vowel pairs, such as /a:/-/a/ and /o:/-/ɔ/, differ in quality as well as duration, though the formants are close together (e.g., Booij, 1995; Adank *et al.*, 2004). There is a general agreement that vowel duration can affect vowel and coda-voicing interpretation in English; there is debate about the consistency and generality of the effects. For instance, Hogan and Rozsypal (1980) showed that in natural speech cues such as vowel, voice bar and frication noise duration are weighted differently in the perception of coda voicing depending on vowel and consonant type in a two-alternative forced-choice task. Hillenbrand *et al.* (2000) showed that both shortened and lengthened vowel duration had a relatively small effect (0%–2%) on interpretation for most vowel pairs but a much larger effect (9%–43%) on others: primarily, shifting among /ɔ, ɑ, ʌ/ and among /ɛ, æ/. The authors accounted for this variation by modeling the similarity among vowel pairs using the dataset of Hillenbrand *et al.* (1995). Dietrich *et al.* (2007) found that Dutch but not English 18-month-olds could learn to associate two objects with two words with /a/ or /æ/ vowels differing only in vowel duration (e.g., [tam]; [t:m]). It is unclear whether adults would show similar cross-language differences. There are almost no available experimental data on how Dutch listeners interpret vowel duration differences without the corresponding differences in vowel quality. Nooteboom and Doodeman (1980) investigated the boundary between /a: / and /a / in a binary forced-choice task and found that for the word pair /tak/-/ta:k / (“branch”—“task”) Dutch listeners perceived artificially shortened long /a: / vowels as /a /, but not lengthened short /a / vowels as /a: /.

In the current study, we presented Dutch and English listeners with native-language words in a transcription task. Unlike most prior studies, we presented listeners with monosyllables in a more “natural” context, with different vowels and varying consonant contexts in a carrier phrase. As described below, half of the test words were presented with natural vowel duration and half with shorter or longer duration. If Dutch and English listeners are equally attuned to durational differences, misperception of vowel duration in Dutch could lead to misperception of vowel identity, whereas in English it could lead to misperception of vowel identity and coda consonant voicing. However, if Dutch but not English listeners represent a duration contrast in their phonological system, they may weigh the duration cue more heavily and be more likely to misperceive vowels based solely on changes in duration.

2. Method

2.1 Subjects

Sixteen native monolingual Dutch listeners (all speakers of Northern Standard Dutch as described in Adank *et al.*, 2004) and 32 native monolingual American English listeners were tested. Participants were undergraduate students at the Radboud University, Nijmegen, The Netherlands, and the University of Pennsylvania, Philadelphia, and received academic credit for their participation.

2.2 Stimuli

Both Dutch and English participants were presented with shortened long vowels and lengthened short vowels, as well as normal-length vowels. In each language, eight different vowels were each presented in six different real words. In addition, six filler non-words were added. The English vowels were the prototypically long /a:, æ, ou, i:/ and prototypically short /ʌ, ɛ, ɪ, u/ (though /u/ is harder to classify; Hillenbrand *et al.*, 2000). The Dutch long vowels were /a:, o:/, the long /ɪ:/ occurring in a pre-/t/ context (orthographically

Table 1. Examples of test words for each vowel.

English						Dutch			
<i>Long</i>	<i>Coda</i>	<i>Example</i>	<i>Short</i>	<i>Coda</i>	<i>Example</i>	<i>Long</i>	<i>Example</i>	<i>Short</i>	<i>Example</i>
a:	voiceless	sock	ɛ	voiceless	met	a:	taak	ɑ	rat
	voiced	cob		voiced	bed		(“task”)		(“rat”)
æ	voiceless	rack	ɪ	voiceless	hit bid	o:	boot	ɔ	pop
	voiced	bad		voiced			(“boat”)		(“doll”)
oʊ	voiceless	boat	ʌ	voiceless	cup bug	ɪ:	peer	ɛ	pet
	voiced	load		voiced			(“pear”)		(“cap”)
i:	voiceless	leak	u:	voiceless	loop	i	diep	ɪ	kip
	voiced	leave		voiced	rude		(“deep”)		(“chicken”)

“eer”), and /i/ (a vowel also characterized as “half-long” (e.g., Booij, 1995); the short vowels were /a, ɔ, ɛ, ɪ/. Examples of test words are listed in Table 1. (Note that some Americans produce /ɔ/ rather than [ɑ] before /g/ in words like *dog*, which could have affected interpretation of this item. Our speaker’s vowel in *dog* was similar to her vowel in other /a/ words, though she does not merge, e.g., *cot* and *caught*. Listeners made no errors on *dog*.) In English, each vowel occurred in three words ending in a voiced coda consonant and three words ending in a voiceless coda consonant. Coda voicing was not a factor in Dutch because of coda devoicing. Because of this extra factor, we doubled the number of participants in the English experiment. Most test words had a CVC structure (Consonant-Vowel-Consonant), apart from 5 English CCVC words and 3 Dutch CVCC words. Most final consonants were stops, apart from 6 Dutch test words ending in the fricative /s/, one in a nasal /n/, 9 in the uvular trill [ʀ], and 4 ending in [ʁ]; 5 English test words ended in fricatives (2 in /s/, 2 in /z/, 1 in /v/). Test words were selected so that most had a real word short/long counterpart, and, for English, a voiced-/voiceless-coda counterpart. For example, the item “bit” with lengthened duration could in principle be misinterpreted as “beat” (attribution of vowel lengthening as signaling a tense vowel), “bid” (attribution of lengthening as signaling a voiced coda), or “bead” (both attributions). Fillers were added to make sure participants reported their interpretations even if they were non-words.¹

Test items for both languages were digitally recorded (44.1 kHz) in the same sound-attenuated room by a native speaker of Dutch and a native speaker of American English using an Audio-Technica (Tokyo, Japan) MB4000C microphone. All items were produced in the utterance-final context “*The next word is __*” or the Dutch counterpart “*Het volgende woord is __.*” Duration and formant frequencies (F1, F2) of all recorded vowels were measured using PRAAT (Boersma and Weenink, 2010). The recordings showed an average $1.92 \times$ ratio between short and long vowels across all recordings for the Dutch speaker and an average $1.78 \times$ ratio between vowels before voiceless and vowels before voiced codas (across tense and lax vowels) for the English speaker. To create the altered-duration vowels, the nuclei of naturally produced vowels were lengthened or shortened with a $1.8 \times$ ratio between long and short vowels, using PSOLA resynthesis in PRAAT. The altered words were screened for naturalness and to confirm that there were no audible discontinuities in the signal. In addition, all words were simply pronounced by our (well-practiced) speakers using a lengthened or shortened vowel. Several tokens of each word were recorded. The natural mispronunciations used in the experiment were selected to match the target vowels in terms of formant frequencies and the duration of the long/short counterpart of the target vowel. Artificial stimuli of the normal-duration vowels were created via a similar manipulation (taking a token pronounced by our speaker using a vowel that was intentionally too long or short and obtaining its appropriate duration via digital manipulation). All English recordings were analyzed to ensure that the coda consonant was always released and produced with appropriate voicing. In Dutch, short vowels were altered to long and vice versa. In English, vowels preceding voiced codas were shortened and vowels preceding unvoiced codas were lengthened.

For this reason, in English each vowel was lengthened and shortened. Each participant heard only one instance of each test word. The condition in which each test word was played (correct/incorrect duration \times artificial/natural duration) was counterbalanced between subjects. Stimuli with non-prototypical vowel durations provided conflicting cues to word identity: Lengthened vowels may have sounded like their phonologically long counterparts and shortened vowels like their short counterparts; and in English only, non-prototypical vowel durations were inconsistent with other cues to coda voicing. Examples of stimuli are available online.

Mm. 1. Dutch stimulus example “lot” and “koop,” with correct (natural and artificial) and altered (natural and artificial) vowels durations. This is a file of type “wav” (1.7 Mb).

Mm. 2. English stimulus example “hit” and “sad,” with correct (natural and artificial) and altered (natural and artificial) vowels durations. This is a file of type “wav” (1.5 Mb).

2.3 Procedure

Participants were instructed to listen to recordings of some real nonexistent words and type in the word that they heard. They were told that we were testing the clarity of the recordings. Participants were tested in a sound-attenuated room in front of a computer screen, wearing Sennheiser (Wennebostel, Germany) HD-465 headphones. On each trial, participants pressed a button labeled “play sound,” which triggered presentation of a spoken sentence. After typing in the perceived word, participants clicked on a button to proceed to the next trial.

3. Results

When words were realized with canonical vowel duration, errors (defined as deviations from the canonical transcription’s vowel or coda voicing value) were rare: 6 instances in Dutch and 2 in English (“knot” heard as *naught* and “peck” as *pack*). We therefore concentrate on errors made on mispronounced stimuli. Overall, Dutch listeners were more strongly influenced by vowel duration (mean, 29.4% of trials, $SD = 7.8$) than English listeners (9.5%, $SD = 5.9$; $t(46) = 9.86$, $p < 0.00001$). Figure 1 shows the distribution of error rates over subjects.

Out of a total of 766 English test trials (32 participants \times 24 trials, minus 2 trials in which participants failed to enter a word) English listeners made only 16 coda voicing errors (2.1%), with 1 participant making three such errors and all others making zero or one. Three of the 48 English target words accounted for 14 of these 16 errors: *leave* ($n = 5$), *loose* (4), and *loop* (5). Only 5 English test items ended in a fricative; thus, out of 80 trials on which incorrect vowel duration was combined with a fricative coda, 11.25% (9 trials) of test words were misperceived. We attribute this tendency to the fact that final-fricative voicing in English is more dependent on the preceding vowel duration cue than final-stop voicing (Broersma, 2009; Hogan and Roszypal, 1980; Raphael, 1972),

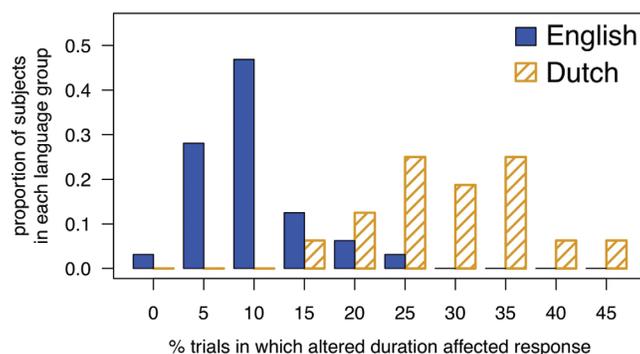


Fig. 1. (Color online) Proportions of Dutch and English subjects against rates of non-canonical interpretation when hearing altered-duration words (dark bars, English; striped bars, Dutch).

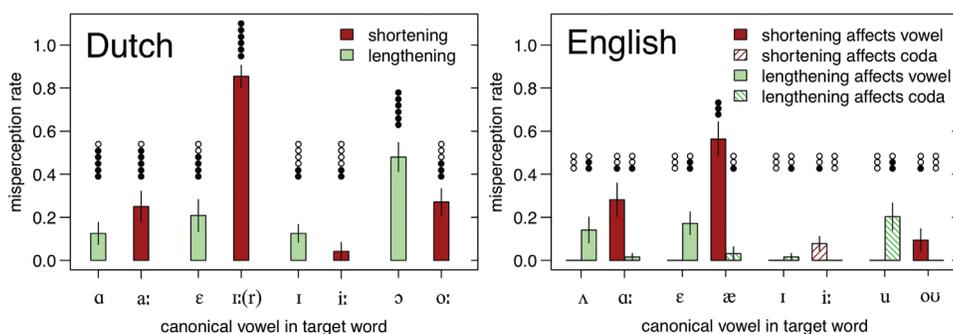


Fig. 2. (Color online) Rates of misperceived vowels in words with altered vowel durations. Dark filled bars give mean vowel error rates for shortened vowels; light bars for lengthened. The English plot (right panel) shows coda voicing errors as hashed bars. Error bars show standard errors over subjects. Filled black circles show the number of different words out of 6 (Dutch) or 3 (English) for which listeners made at least one error.

and voiced final fricatives are very often produced without actual voicing (e.g., Haggard, 1978). English listeners made 57 vowel identity errors (7.4%).

Examination of the results vowel by vowel shows that duration-induced changes of interpretation did not occur uniformly. Figure 2 (left panel) shows the Dutch error rates per vowel. All errors were of the predicted sort, namely short-vowel interpretations when long vowels were shortened (dark bars) and long-vowel interpretations when short vowels were elongated (light bars). From vowel to vowel, errors were largely consistent across words: Of the 48 words, 35 resulted in at least one misperception. By contrast, in English only 15 of the 48 words yielded a misperception (Fig. 2, right panel). The proportions 15/48 and 35/48 differ significantly (proportion test, $X(1) = 15.1$, $p < 0.001$). The most frequent of the Dutch misperceptions was in /ɪ:r/; we speculate that /r/-coloring of vowels leads listeners to rely less on the vowel's spectral features (and more on duration). Among English listeners, the most common errors emerged in exchanges between /ɛ/ and /æ/, which are relatively similar in spectral characteristics (e.g., Hogan and Rozsypal, 1980). The fewest errors emerged on the spectrally less similar /i:/ - /ɪ/ pair, matching the results of Hillenbrand *et al.* (2000).

Further analysis contrasted Dutch and English listeners' vowel interpretation errors using a series of mixed logit models. The models included *subjects* and *targets* (within *vowels*) as random effects and introduced *language*, *manipulation* (shortening vs lengthening), *artificiality* (natural vs artificial duration changes), and the interactions of these factors as fixed effects. Word frequency was modeled using the difference in Dutch Center for Lexical Information (CELEX) log frequency (per million) between the target word and the word resulting from vowel mispronunciation (Baayen *et al.*, 1995). Predictors were retained in the final model if they improved the model's fit. Statistics of the best-fitting model are given in Table 2.

Table 2. Summary of fixed effects in mixed logit model ($N = 1152$ observations; log-likelihood = -288.5) with *language* (reference *Dutch*), *manipulation* direction (reference *lengthen*), and *artificiality* (reference *natural*) as binary predictors of misperception, as well as *frequency*, the log frequency (per million) of the target word minus that of the vowel-change competitor. Sum contrast weights are given in parentheses.

Predictor	Coeff.	Std. Err.	Z	P
Intercept	-2.777	0.505	-5.50	<0.0001
Frequency [target minus vowel competitor]	-0.941	0.236	-3.99	<0.0001
Language (English 0.5, Dutch -0.5)	-3.573	1.012	-3.53	<0.0005
Manipulation (shorten 0.5, lengthen -0.5)	1.954	0.752	2.60	<0.01
Artificiality (artificial 0.5, natural -0.5)	0.133	0.256	0.52	ns
Language \times artificiality	2.786	0.563	4.95	<0.0001
Artificiality \times manipulation	-2.080	0.554	-3.75	<0.0002

The analysis confirmed the greater error rate among Dutch listeners. In addition, errors were more likely when vowels were shortened than when lengthened. English listeners tended to make more errors on digitally altered vowels, while Dutch listeners made more errors on naturally altered vowels. If Dutch listeners consider duration phonemic, they may be attuned to (Dutch) phonetic correlates of duration, which were present in natural but not artificial mispronunciations; by contrast, artificial manipulation may have simply increased noise and uncertainty for English listeners. A second unanticipated effect was an interaction between *manipulation* and *artificiality* (but not *language*): the prevalence of shortening-induced errors was stronger for naturally mispronounced words. In sum, although the specifics of our implementation of the mispronunciations had some complex effects, Dutch listeners were more strongly affected by durational changes than English listeners for both implementations.

4. Discussion

Our findings show that in citation-form speech, alterations of vowel duration affect Dutch listeners' interpretation more than they affect English listeners. In this sense the results align with previous claims that Dutch has a phonological duration contrast and thereby differs from English. We note though, that even in Dutch, the long/short opposition is not controlled entirely by duration; for most vowels, the duration manipulation affected interpretation on less than 50% of trials. Our English results are broadly consistent with previous findings. Nooteboom and Doodeman (1980), testing the /a:/ - /a/ vowel pair, found errors provoked by shortening but not lengthening. Our results for this pair are in the same direction, but this asymmetry was not consistently observed across all long/short vowel pairs. Lengthening might show weaker effects because it is available in the language for prosodic effects such as application of emphatic stress (Ko *et al.*, 2009). If short vowels are lengthened more often than long vowels are shortened, a perceptual asymmetry could result. Also, lengthening may facilitate perceptual access to vowel quality, whereas in shortened vowels vowel quality may be harder to evaluate (leading to reliance on duration).

5. Conclusion

Dutch and English are similar in terms of syllable structure, variability in lexical stress with optional vowel reduction in unstressed syllables, and opposition of long/short vowels with correlated quality and durational differences in canonically realized vowels. Yet when hearing clearly articulated words, Dutch listeners were more strongly affected by manipulation of vowel duration than English listeners. Thus, similar acoustic cues are implemented differently in the perceptual systems of Dutch and English listeners.

Acknowledgments

An NWO (Dutch National Science Foundation) Rubicon Grant to S.V.H.v.d.F. and National Institutes of Health (NIH) grant R01-HD049681 to D.S. supported this research. We thank Delphine Dahan, Paula Fikkert, Jane Park, Allison Britt, Josef Fruehwald, and all participants for their help with this study.

References and links

¹See supplemental material at <http://dx.doi.org/10.1121/1.3532050> Document No. E-JASMAN-129-502102 for a table listing item characteristics and results for each item. For more information see <http://www.aisp.org/pubservs/epaps.html>.

Adan, P., Van Hout, R., and Smits, R. (2004). "An acoustic description of the vowels of Northern and Southern Standard Dutch," *J. Acoust. Soc. Am.* **116**, 1729–1738.

Baayen, R. H., Piepenbrock, R., and Gulikers, L. (1995). The CELEX lexical database (Release 2) [CD-ROM]. (Linguistic Data Consortium, University of Pennsylvania, Philadelphia).

Boersma, P., and Weenink, D. (2010). PRAAT: Doing Phonetics by Computer, Version 5.1.32 available at <http://www.praat.org/> (Last viewed July 30, 2010).

Booij, G. (1995). *The Phonology of Dutch* (Oxford University Press, Oxford).

- Broersma, M. (2009). "Perception of final fricative voicing: Native and nonnative listeners' use of vowel duration," *J. Acoust. Soc. Am.* **127**, 1636–1644.
- Chomsky, N., and Halle, M. (1968). *The Sound Pattern of English* (Harper and Row, New York).
- Denes, P. (1955). "Effect of duration on the perception of voicing," *J. Acoust. Soc. Am.* **27**, 761–764.
- Dietrich, C., Swingley, D., and Werker, J. F. (2007). "Native language governs interpretation of salient speech sound differences at 18 months," *Proc. Natl. Acad. Sci. U.S.A.* **104**, 454–464.
- Haggard, M. (1978). "The devoicing of voiced fricatives," *J. Phonetics* **6**, 95–102.
- Hillenbrand, J. M., Getty, L. A., Clark, M. J., and Wheeler, K. (1995). "Acoustic characteristics of American English vowels," *J. Acoust. Soc. Am.* **97**, 3099–3111.
- Hillenbrand, J. M., Clark, M. J., and Houde, R. A. (2000). "Some effects of duration on vowel recognition," *J. Acoust. Soc. Am.* **108**, 3013–3022.
- Hogan, J. T., and Rozsypal, A. J. (1980). "Evaluation of vowel duration as a cue for the voicing distinction in the following word-final consonant," *J. Acoust. Soc. Am.* **67**, 1764–1771.
- House, A. S. (1961). "On vowel duration in English," *J. Acoust. Soc. Am.* **33**, 1174–1178.
- Klatt, D. H. (1976). "Linguistic uses of segmental duration in English: Acoustic and perceptual evidence," *J. Acoust. Soc. Am.* **87**, 820–857.
- Ko, E., Soderstrom, M., and Morgan, J. (2009). "Development of perceptual sensitivity to extrinsic vowel duration in infants learning American English," *J. Acoust. Soc. Am.* **126**, EL135–EL139.
- Moulton, W. G. (1962). "The vowels of Dutch: phonetic and distributional classes," *Lingua*, 294–312.
- Nooteboom, S. G., and Doodeman, G. J. N. (1980). "Production and perception of vowel length in spoken sentences," *J. Acoust. Soc. Am.* **67**, 276–287.
- Raphael, L. J. (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.