Perceived vowel duration

Carlos Gussenhoven

Radboud University Nijmegen

1 Introduction

Two subjects that lie close to Sieb Nooteboom's professional heart are speech perception and vowel duration, and I am therefore pleased that I can combine these two interests in my contribution to this volume. I am concerned with the difference between *acoustic duration* and *perceived duration*, which concepts differ in a similar way to fundamental frequency and pitch. I will claim that vowel height affects perceived duration, in the sense that higher vowels sound longer than lower vowels when acoustic durations are equal. That is, vowel height and perceived duration are positively correlated. In section 2, I present the results of a perception experiment with Dutch listeners which shows this correlation. In sections 3 and 4 I deal with the two main questions that this finding raises. The first concerns the reason for this correlation. I will argue that it is to be sought in a mechanism of compensatory listening, and will cite other cases in the literature that have been given parallel explanations. The second question is whether the correlation is of any significance for the phonetics or phonology of languages. Here, I will argue that it solves a two cases of vowel raising, one phonetic and one phonological, in English and Limburgian Dutch, respectively.

2 Perceived duration of high and mid vowels

The research reported in this section was carried out in collaboration with Wilske Driessen, who wrote her MA thesis on this topic (Driessen 2004). A female speaker of Dutch recorded a number of isolated pronunciations of the vowels [i, y, u, ɛ, œ, ɪ, ɔ] as in *wie* 'who', *nu* 'now', *koe* 'cow', *bed* 'bed', *oeuvre* 'works', *pit* 'kernel', *bot* 'blunt', respectively, on digital audiotape, pronouncing them with a weakly falling intonation. Good tokens of these six vowels were stored at a 16 kHz sampling rate and trimmed so as to end up with a complete number of periods that were the closest to 180 ms in duration. With the help of the option for the manipulation of the fundamental frequency in the Praat package (Boersma & Weenink, 2002), they were provided with contours starting at 160 Hz and ending at 110 Hz, with a 40 ms plateau of 220 Hz beginning after 40 ms. Each of these standardized speech files then served as the basis for six further manipulated vowels with 15 ms increments in duration. That is, we ended up with seven durational versions of each vowel: 180, 195, 210, 225, 240, 255, and 270 ms. The same treatment was applied to tokens of four further vowels, [a] as in *na* 'after' and the diphthongs [ɛi, œy, ʌu], as in *ei* 'egg', *ui* 'onion', *kou* 'cold'. The diphthongs served as fillers for the purposes of this report, but the result of [a] will be reported here. The set of 10×7 stimuli was randomised three times and divided into blocks of ten. Thirty-four Dutch listeners rated each stimulus for duration on a 7-point scale, with the shortest duration appearing on the left of the scale. Each block was preceded by an anchor stimulus of 225 ms with a schwa-like vowel quality, which corresponding to a scale on the answer sheet in which the fourth scale category had been crossed off. Listeners were told that this stimulus represented the mid-point on the scale.

An analysis of variance with Duration (7 levels) and Vowel Height (2 levels) showed that acoustic duration and vowel height significantly affected the perceived duration.

Figure 1 shows that the effect of acoustic duration is consistently present in the case of all six vowels, while [a], which had been excluded from the analysis, shows the same positive correlation. More interesting in the context of this contribution is the finding that the high vowels are consistently rated longer than the equivalent mid vowels [$F(1,32)=18.11$, $p<.01$]. Interestingly, mid back [ɔ] is rated as longer than low back [a]. The results for the high and mid vowels have meanwhile been replicated in a second experiment with different stimuli and a different group of Dutch listeners.

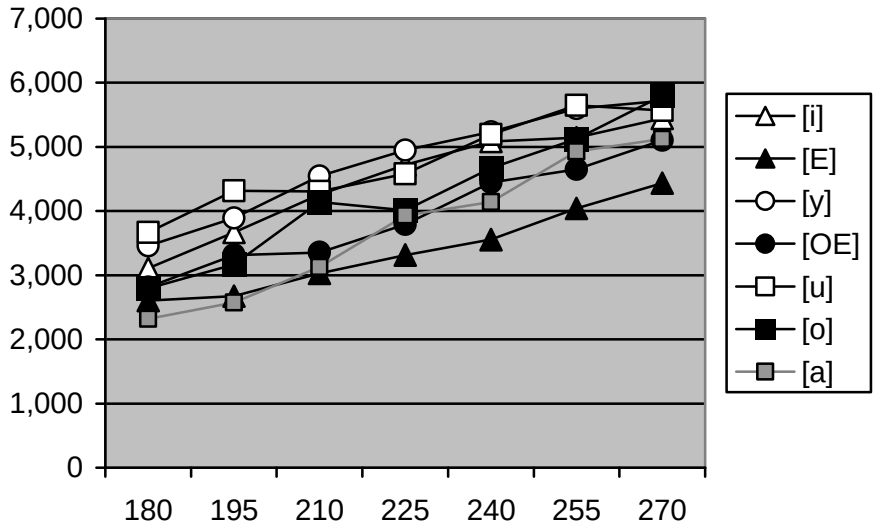

Figure 1. Perceived durations (on a 7-point scale) by Dutch hearers of seven vowel stimuli, each with seven acoustic durations (in ms) .

3 The explanation of the correlation between vowel height and perceived duration

Higher vowels are shorter than lower vowels. This is a universal tendency which has been explained on the basis of the distance between the roof of the mouth and the articulatory excursion of the tongue(-cum-jaw) made for the vowel: the greater this distance, the longer the vowel. In Dutch, this tendency has become phonologized. Originally long /i, y, u/, as in *wie* 'who', *nu* 'now', *koe* 'cow', have become short /i, y, u/, thus merging their quantity with /ɪ,ʏ/, as in *pit* 'kernel' and *put* 'sink, well' (Moulton, 1962; Nooteboom, 1972; Gussenhoven, 2004). It is suggested that, paradoxically, the negative correlation between vowel height and acoustic duration explains why vowel height and *perceived* duration are *positively* correlated. The hearer knows that low vowels require more time and thus be inherently longer than high vowels. When assessing the duration of a vowel he will therefore subtract this inherent portion in the duration before constructing the perceived duration. Putting it differently, high vowels do not, but low vowels do include a component in the *articulatory duration* which is obtained as an unintended bonus for producing the vowel *quality* in question, and by way of compensation, the hearer reduces the acoustic duration accordingly.

The explanation readily generalizes to other cases of 'compensatory listening'. First, Pierrehumbert (1979) found that accent-lending fundamental frequency peaks in English have more prominence if they come later in the utterance. The effect was attributed to the existence of a descending, abstract reference line marking equal pitch, which mimicked the declination found in production studies. By employing this reference line instead of the fundamental frequency scale when measuring the pitch of the peak, the hearer thus compensated for the declination in production, by bringing late low peaks back up to the level it would have had if there had been no declination. A second case is Silverman (1984), who demonstrated that the pitch of the same fundamental frequency peak was lower when combined with the British English vowel /iː/ than when it combined with the vowel /ɑː/. He did this by having hearers judge the pitch of the accented words in utterances like *They only FAST before FEASTing* and *They only FEAST before FASTing*, in which the only difference is the spectral composition of the accented words, and demonstrating that *feast-* had lower pitch than *fast-* when fundamental frequency contours were identical.

Both the declination effect and the intrinsic pitch effect occur because the hearer subtracts the effect of an articulatory advantage from the acoustic value. In the first case, the high subglottal pressure affords the speaker an advantage in the fundamental frequency domain, since higher subglottal pressure will lead to increased rates of vocal fold vibration. In the second case, the articulation of the high vowel causes an upward pull on the tongue root, the hyoid and the thyroid, causing some tensing of the vocal folds, which as a result vibrate a little faster (see Maddieson, 1997). Figure 2 presents this idea of subtraction of an articulatory advantage in a graphic form. The shaded area represents the boost in fundamental frequency due to high subglottal pressure in the case of the declination effect, the boost in fundamental frequency due to high tongue position in the case of the 'intrinsic pitch' affect, and lastly, the boost in vowel duration due to low tongue(-cum-jaw) position in the case of the correlation which is the subject of this contribution.

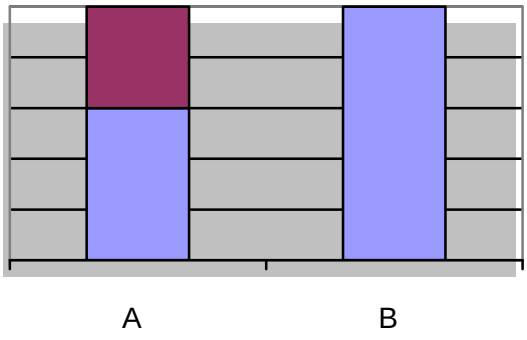

Figure 2. Compensatory listening explained as the hearer's subtraction of an articulatory advantage enjoyed by the speaker. Sounds A and B have identical acoustic values, but the hearer reduces A's value by subtracting the boost of the fundamental frequency or duration, which he assumes occurred as a by-product of the articulation.

4 The correlation between vowel height and perceived duration in phonetics and phonology

It is unlikely that the main finding reported in Figure 2 is a psycho-acoustic effect, unrelated to speech. This is because its explanation generalizes to other correlations that have been reported. Clearly, hearers bring detailed phonetic knowledge of speech production to bear on

the interpretation of the acoustic data. Not only are speakers in control of their speech production, hearers too know how speakers go about producing speech. By itself, of course, this merely shows that the correlation between vowel height and perceived duration is known to the participants in speech communication, but it has not been shown that it has any significance other than the significance which a known underwater rock at the entrance of a harbour has for a wary ship captain. The purpose of this section is to suggest that there is more to it. Section 4.1 summarizes certain regularities in pre-obstruent vowels in English and and an explanatory account of them by Moreton (2004), and introduces similar facts that are used to enhance a tonal contrast in Limburgian Dutch. Section 4.2 proposes an alternative explanation in which the observed correlation between vowel height and perceived duration plays a crucial role.

4.1 Enhancing the laryngeal contrast in English syllable codas

In a recent article, Moreton (2004) summarizes a number of research findings showing that in English low vowels are opener before [–voice] obstruents than before [+voice] obstruents. That is, *cat, boss* have a higher F_1 than *cad, Bozz*. Moreton assumes that this is an effect of hyperarticulation, and bases this assumption on the even more widely reported finding that the second element of English closing diphthongs is higher before [-voice] obstruents than before [+voice] obstruents. The latter effect has been phonologized in the case of /aɪ/ in Canadian English, where *writer*, for instance, has a categorically different vowel from *rider*, and where the distribution of these vowels is not entirely predictable (Wells, 1981:495). Although I will argue that Moreton's conclusion is incorrect, the fact that low vowels are lower and high second elements of diphthongs are higher before [–voice] indeed suggests hyperarticulation of the vowel. Moreton reports new measurements showing that the first elements of diphthongs are also higher before [–voice] consonants, though not to the same extent as second elements, and additionally demonstrates that F_1 and F_2 in the second elements of diphthongs may influence the perception of [voice] in the following consonant. After presenting and convincingly rejecting a number of explanations that have been proposed in the literature for these findings (not all of which are based on the assumption that the difference results from hyperarticulation), he tentatively advances the explanation that the vowels are hyperarticulated because the voiceless consonants are hyperarticulated, as if by a leakage of articulatory effort preceding the consonant in question. Since the greatest difference with the pre-voiced context is achieved in the latter half of the vowel, also in the case of the low monophthongs, this explanation would at first sight seem to be on the right track.

There are two problems with Moreton's suggestion of his *Spread-of-Facilitation* hypothesis. The first concerns the status of hyperarticulation as way of enhancing contrasts. It is not clear why hyperarticulation should apply to only one of two members of an opposition, rather than to both. Related to this is the question why, if one is to be chosen, this should be the voiceless member. Moreton suggests that the answer to this question is that the closing gesture of the voiceless obstruent is what is specifically used to enhance the contrast, rather than the opening gesture, and that therefore the hyperarticulation is only found before the voiceless consonant. This explanation is not unreasonable; yet, as far as I know it introduces a new conception of contrast enhancement, *viz*. that of selectively hyperarticulating one member of an opposition instead of both, and thus using the hyperarticulation itself as creating a contrast with the non-enhanced member of the opposition. The second problem with Moreton's conjecture is that there is no evidence that high vowels are higher before [-voice] obstruents than before [+voice] obstruents. As he observes, this is one of the predictions that his theory

makes. Like Wolff (1978) and Van Summers (1987), Moreton investigated the behaviour of *low* vowels before the voicing contrast.

While the first arguments against the *Spread-of-Facilitation* hypothesis might be countered with an observation that apparently this is the way things go, and the second by observing that there the jury is still out as long as the data are not available, there is a third problem which cannot as easily be dismissed. Vowel quality differences similar to those that have been observed before the laryngeal contrast in English have been found in syllables with a tone contrast in Dutch (Limburgian) dialects spoken in the northeast of Belgium and the southeast of the Netherlands (Gussenhoven, 2004). The tone contrast is known as Accent 1 vs Accent 2, or in the Dutch dialectological literature as *stoottoon* or *valtoon* ('pushing tone/falling tone') vs *sleeptoon* ('dragging tone'), respectively. Phonologically, the contrast has been described as the absence of a lexical tone vs the presence of H, respectively (Gussenhoven & Aarts, 1999). Phonetic realizations vary across the dialects, but frequently reported differences are that Accent 1 tends to be shorter, to have wider F_0 movements, and, less systematically, to have a firm amplitude decrease towards the end of the sonorant segments in the rhyme. For instance, in the dialect of Mechelen-aan-de-Maas, mid vowels split into an opener and closer vowel in syllables with Accent 1 and Accent 2, respectively, as shown in (1) (Verstegen, 1996).

(1)

Accent 1		*Accent 2*	
ɣɛɛl	'yellow-ATTR'	ɣeel	'yellow-PRED'
wɛɛx	'road-PL'	weex	'road-SG'
ɣɔɔn	'go-1SG,PRES'	ɣoon	'go-1PL,PRES'
nɔɔl	'needle-SG'	noolə	'needle-PL'

And in the dialect of Maastricht, the diphthongs /ɛi, œy, ɔu/ have markedly different allophones depending on whether they cooccur with Accent 1, as in (2a), or Accent 2, as in (2b) (Gussenhoven & Aarts, 1999). When combining with Accent 1, the diphthong's end point is very close, while in syllables with Accent 2 the end point is only weakly approximated, so much so that these vowels may lose their diphthongal character. The difference is non-discrete, and native speakers regard the allophones as the same vowel in each of the three cases.

(2) a. Accent 1: / bɛi/ 'bee', /lœy/ 'people', /dɔuf / 'pigeon'; [bɛj, lœj, dɔwf]

b. Accent 2: /bɛi/ 'near', /lœy/ 'lazy', /dɔuf / 'deaf'; [ˉbɛː$^{(i)}$, ˉlœː$^{(y)}$, ˉdɔː$^{(u)}$f]

Strikingly, the closer second elements of the diphthongs and the opener realizations of the monophthongs go hand in hand, as in the cases reviewed by Moreton. However, equally strikingly, the contrast that is to be enhanced by these quality differences is not a laryngeal contrast, but a tonal one.

4.2 An explanation for English and Limburgian Dutch

Before continuing to speculate on the explanation for the quality differences, we need to consider the question what the English coda voicing contrast and the Limburg tone contrast have in common. The answer is that both contrasts are enhanced, in Stevens & Keyser's (1998) sense of aided by some non-primary phonetic parameter, by a durational difference. I therefore propose that a *Duration Enhancement* hypothesis should take the place of Moreton's *Spread-of-Facilitation* hypothesis. Specifically, [–voice] obstruents are preceded by shorter sonorant portions in the rhyme than [+voice] obstruents, while also the sonorant portions in rhymes with Accent 1 are shorter than those with Accent 2. Since Limburgian

Dutch Accent 1 patterns with English [–voice] codas in this respect, it is difficult to escape the impression that the vowel quality differences are there to enhance the duration differences.

If this is so, the next question is how the features associated with the shorter vowels, closer second elements of diphthongs and opener low vowels, can contribute to the impression of shorter vowel duration. It would appear that they do so in different ways. The high second element is there to change the second element of the diphthong from a vowel into a consonant, and in that way reduce the perceived vowel duration. That is, [ɛj] sounds as if it has a shorter vowel than [ɛi], which in turn may well sound sorter than [ɛe]. In this interpretation, Canadian Raising is a trick to transfer part of the vowel to a consonantal percept, thereby reducing the perceived vowel duration.

Second, the more open vowels before [–voice] obstruents and in syllables with Accent 2 are likewise there to suggest longer vowel durations, in this case by exploiting the compensatory listening effect reported in section 2. That is, vowel lowering and vowel raising are ways of making vowels sound shorter and longer, respectively. The correlation between vowel height and perceived duration therefore appears to have a crucial role in the phonetics and the phonologies of languages.

5 Conclusion

The hypothesis that vowel lowering and closer diphthong off glides are ways of making vowels sound shorter has a number of advantages over Moreton's *Spread-of-Facilitation* hypothesis. First, it is no longer the case that *one* of the two terms in the phonological contrast to be enhanced is selected for having the privilege of a more canonical articulation bestowed upon it. In the *Duration Enhancement* hypothesis the two terms receive in principle equal treatment. Indeed, the number of phonological repercussions of the durational enhancement in the Limburgian Durch dialects is quite varied, and Accent 1 may just as easily be targeted for a change as Accent 2. Second, the *Duration Enhancement* hypothesis is capable of explaining the vowel quality adjustments in both as an enhancement of the English laryngeal contrast and of the Limburgian Dutch tone contrast.

Importantly, the choice between the two theories can be decided by a simple experiment. First, if the *Spread-of-Facilitation* hypothesis is correct, British English words like *bit*, *niece*, *look*, *belief* should have higher vowels than *bid*, *knees*, *Luke*, *believe*, but if the *Duration Enhancement* hypothesis is correct, the former should have lower vowels than the latter. Second, if the *Spread-of-Facilitation* hypothesis is correct, British English words with mid vowels like *bet*, *Bert*, *boss* should have more peripheral vowels than *bed*, *bird*, *Bozz*, but if the *Duration Enhancement* hypothesis is correct, the former should, again, have lower vowels than the latter. Research that addresses these questions is in progress.

References

Boersma, P. & D. Weenink (2002). Praat: Doing Phonetics by Computer. Computer program, available at http://www.praat.org .

Driessen, W. (2002). Compensatory Listening: The Effect of Vowel Quality on the Perception of Vowel Duration. MA thesis, Department of English, University of Nijmegen.

Gussenhoven, C. & W. Driessen (2004). Explaining two correlations between vowel quality and tone: The duration connection. Paper presented at the ISCA workshop on *Prosody 2004*.

Gussenhoven, C. & F. Aarts (1999). The dialect of Maastricht. *Journal of the International Phonetic Association, 29*, 55-66.

Maddieson, I. (1997). Phonetic universals. In W.J. Hardcastle & J. Laver (Eds.) *The Handbook of Phonetic Sciences* (pp. 619-639). Oxford: Blackwell.

Moreton, E. (2004). Realization of English postvocalic [voice] contrast in F_1 and F_2. *Phonetica, 32*, 1-33.
Moulton, W. (1962). The vowels of Dutch: Phonetic and distributional classes. *Lingua, 11*, 294-312.
Nooteboom, S.G. (1972). *Production and Perception of Vowel Duration: A Study of Durational Properties of Vowels in Dutch*. PhD dissertation, Rijksuniversiteit Utrecht.
Stevens, K. & J. Keyser (1989). Primay features and their enhancement in consonants. *Language, 65*, 81-106.
Van Summers, W. (1987). Effects of stress and final consonant voicing on vowel production: Articulatory and acoustic analysis. *Journal of the Acoustical Society of America, 82*, 847-863.
Verstegen, V., 1996. Bijdrage tot de tonologie van Oostlimburgse dialecten. In H. van de Wijngaard (Ed.) *Een eeuw Limburgse dialectologie* (pp. 229-234). Hasselt/Maastricht: VLDN/Vereniging Veldeke Limburg.
Wells, J.C. (1981) *Accents of English*. Three volumes. London: Longman.
Wolf, C.G. (1978). Voicing cues in English final stops. *Journal of Phonetics, 6*, 299-309.