A vowel height split explained: Compensatory listening and speaker control

Carlos Gussenhoven

Abstract

The raising of long vowels and the lowering of short vowels, a phenomenon observed in a variety of languages, is to be explained as the speaker's capitalization on the compensatory perception of vowel height. Higher vowels sound longer than lower vowels, by way of compensation for the articulation-driven lengthening of open vowels. Speakers may exploit this effect by signalling short duration through vowel lowering and long duration by vowel raising. Two cases are presented in which the durational contrast itself serves to enhance a phonological opposition, a laryngeal opposition in English coda obstruents and a tonal opposition in Limburgian dialects of Dutch. A related correlation in vowel quality, the strengthening of diphthongal off-glides in short contexts versus the monophthongization in long contexts, is likewise to be explained as aiding the perception of the phonetic duration differences, although the effect is based on a different strategy, that of transferring the off-glide to a consonantal glide which is no longer included in the perceived vowel duration. Perceived vowel duration should be as carefully distinguished from acoustic vowel duration as pitch usually is from fundamental frequency. We present the results of three experiments to support these claims.

1. Introduction

1.1. Contrast enhancement

Phonological contrasts may be phonetically realized with the help of other parameters than the 'primary' phonetic feature (Stevens and Keyser 1989). The contribution of an 'enhancing' feature to the distinctiveness of the contrast is often more salient than that of the primary feature. For instance, the durational difference between the sonorant segments in English syllable rhyme before fortis and lenis codas, e.g., *beat, pint* versus *bead, pined*, arguably contributes more to

the distinctiveness of the laryngeal coda contrast than the voicing difference in the obstruents. The focus in this paper is on the rationale for the way in which enhancing features are used to bring out the salience of the primary contrast.

It is assumed that enhancing features, just like primary features, are under the control of the speaker, and are thus language-specific, but that they follow tendencies that are inherent in the process of speech production and perception (Kingston and Diehl 1994). In the example given, the shortening of sonorant portions in a rhyme closed by [-voice] obstruents, henceforth 'pre-fortis clipping' (Wells 1981), is explained as a signal to the listener that the following obstruent is phonetically long. Because the transglottal pressure difference creating the air flow driving vocal fold vibration is hard to maintain in the face of the impedance by the oral constriction of obstruents, [+voice] obstruents tend to be shorter than [-voice] obstruents. This natural difference in duration between the two types of consonants is thrown in relief by the complementary duration difference in the preceding sonorant segments: if the sonorant portion of the rhyme is long, the obstruent will be short, given a tendency to make rhymes equally long (Catford 1977: 197). Enhancing features thus go with the tide, exaggerating the manifestations of ebb and flow.

Two aspects of enhancement will be brought out by the particular cases to be dealt with here. First, enhancement may be indirect in that concomitant aspects of the contrast are enhanced. Pre-fortis clipping does nothing to make the voicelessness of the obstruents in question actually *sound* voiceless, nor does its absence make the vocal fold vibration of [+voice] obstruents stand out. Speakers may thus enhance enhancement features. Second, enhancement may rely on the ability of the listener to make inferences on the basis of his knowledge of speech production and perception. Pre-fortis clipping illustrates this second aspect, in that the speaker's behavior makes no sense without the knowledge that voiced obstruents tend to be short.[1] Enhancement may thus result from a give-and-take between speaker and hearer, whose interests are often in conflict. In this case, the implementation of pre-fortis clipping (shortening of the sonorant portion in one context and lengthening it in the other) is a concession to the hearer by way of compensation for the frequent devoicing of the voiced obstruent.

1.2. The problem

A thus far poorly understood set of phonetic adjustments occurs before [–voice] coda obstruents in English. Moreton (2004) summarizes

research findings showing that English low vowels are *lower* before [–voice] obstruents than before [+voice] obstruents. That is, *cat*, *boss* have a higher F1 than *cad*, *Bozz*. Second, the second elements of English closing diphthongs are *raised* before [–voice] obstruents, as in *rice*, *house*, relative to their pronunciation before [+voice] obstruents, as in *rise, to house*.

1.3. A failed explanation

After disposing of a number of unsatisfactory explanations, Moreton tentatively assumes that pre-fortis vowel lowering and endglide raising are due to hyperarticulation. Since in both cases the vocalic gesture is less central, more peripheral, before [–voice] codas than before [+voice] codas, he assumes that [–voice] codas trigger hyperarticulation of the preceding vowel. He supports this view with new measurements showing that the *first* elements of diphthongs are also higher before [–voice] consonants, but to a lesser extent. The raised realization of these low first elements must therefore be a coarticulatory effect of the closer off-glide. Since the raising of low vowels cannot be interpreted as hyperarticulation, this fact strengthens his explanation. The timing of this brief phase of hyperarticulation, moreover, fits in with an earlier finding that the lowering of low monophthongs appears to be most extreme in the latter portion of the vowel (Van Summers 1987).

If Moreton's account is correct, the question arises why the hyperarticulation should occur where it does. Moreton tentatively answers that question by hypothesizing that the hyperarticulation occurs as if by a leakage of the articulatory effort expended on the following fortis consonant. This hypothesis is summarized as in (1).

(1) *Spread-of-Facilitation Hypothesi*s (Moreton 2004):
Low monophthongs and high off-glides are hyperarticulated before [–voiced] obstruents because contrastively [–voice] obstruents are hyperarticulated.

An *a priori* problem with (1) 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. Moreover, given that one is to be chosen, it is not clear why this should be the voiceless member, and why the leakage of hyperarticulation occurs *before* rather than after the closure. With respect to this last point, Moreton suggests that the closing gesture of the [–voice] obstruent is the most energetic element in the opposition,

rather than the opening gesture or than either the opening or closing gesture of the [+voice] obstruent. What remains uncomfortable, however, is the notion of hyperarticulation as a way of creating a contrast with a non-hyperarticulated member of an opposition.

There are two ways in which (1) can be put to the test. A strong prediction is that, because they are hyperarticulated, *high monophthongs* will be *higher* before [–voice] obstruents. In section 2, it will be shown that this prediction is incorrect. The second way is to look for similar vowel adjustments in other languages, and see whether (1) explains these other cases too. Section 3 presents a closely parallel set of vowel adjustments which are evidently unrelated to a laryngeal contrast, and instead co-occur with the members of a tone contrast, in dialects of Dutch spoken in Belgium and the Netherlands, henceforth Limburgian. In section 4, I present a new explanation, which applies both to the English and the Limburgian data. It is based on the assumption that the vowel adjustments exist because they enhance the duration differences which are used in both languages as enhancements of a laryngeal coda contrast and a tone contrast, respectively. With the help of two perception experiments, the question is subsequently answered why these particular vowel adjustments, vowel lowering and endglide raising, should be able to help vowels to sound shorter.

2. The behavior of English high vowels before [–voice] codas

Like Wolff (1978) and Van Summers (1987), Moreton investigated the behavior of *low* monophthongs before the voicing contrast. Hypothesis (1) of course equally makes a prediction about the high monophthongs /iː, ɪ, uː, ʊ/. Specifically, these high vowels should have lower F1 before [–voice] obstruents, reflecting their predicted raised, i.e., hyperarticulated pronunciation. The prediction with respect to the F2 of these vowels before fortis codas is less homogeneous; it might be higher in the case /iː, ɪ/ and lower in the case of /uː, ʊ/, acoustic differences that will result from more peripheral tongue positions. Hillenbrand, Clark and Nearey (2001) report the effects of different types of preceding and following consonants on F1, F2 and F3 of four high and four low English vowels as spoken by twelve speakers, and report higher F1 for vowels following as well as preceding [–voice] plosives, indicating a slight raising in voiced surroundings.[2] An inspection of their Fig. 11 reveals that the effect is stronger for low vowels than for high vowels. More recently, Hawkins and Nguyen (2004) similarly found lower F1 across a range of vowel heights before [–voice] than before [+voice] codas, as part of a complex of syllable-

wide phonetic differences they interpret as representing a 'sombre' vs. 'bright' contrast attending the [+voice]-[–voice] coda contrast. Since those investigations were not carried out so as to put Moreton's hypothesis to the test, a production experiment was conducted involving the English high vowels /iː, ɪ, uː, ʊ/. If high vowels indeed turn out to be raised before [+voiced] coda obstruents, this would seriously undermine the hyperarticulation hypothesis in (1).

Two male native speakers of English, one aged 36 from Texas and mildly Texas-accented (TP), and one aged 58 from the south of England and speaking with an RP accent (EK), produced two DAT-recordings of a corpus of brief sentences in which monosyllabic words appeared in phrase-final position. Half of these words ended in /t,s/ and half in /z,d/. These words were *seat, geese; kit, bliss; suit, loose; foot; seed, keys; kid, Liz; sued, snooze; stood.* In the corpus, each word appeared in a semantically appropriate sentence, e.g. *There's still an empty seat*, as well as in both positions in the frame *Not --, but ---*, where each word with a [–voice] coda was contrasted with one containing the same vowel but ending in a [+voice] coda, e.g. *Not seat, but seed*. In the case of /iː, ɪ, uː/ this procedure yielded six tokens of each vowel-consonant combination, or 12 vowel-coda type combinations after merging the data for fricatives and plosives. In the case of /ʊ/, for which we had no words ending in a coronal fricative, we obtained a total of eight tokens per coda type by adding a sentence contrasting *foot* and *good*. The speakers were allowed to repeat any sentence as often as they wished, and the last intended utterance was selected as the speech file in which speech sections were selected which corresponded to the experimental vowel, from the first reliable period up to and including the last. Subsequently, the duration of the vowels as well as the F_0, F1 and F2 at 25%, 50% and 75% of the vowel duration were measured.

In Fig. 1, mean F1 and F2 at the 50% point of the tense vowels /iː, uː/ (panels a and b) and the lax vowels /ɪ, ʊ/ (panels c and d) are plotted as a function of coda type, for the two speakers separately. The length mark indicates the value of each vowel in the [+voice] context. As can be seen, in all eight comparisons the vowel before the [–voice] obstruents is lower than the same vowel before the [+voice] obstruents. The question whether coda type influences the degree of opening of the vowel can be answered in our data by looking at the F1 at the 50% point of the vowel. Analyses of Variance were performed for each speaker separately with VOWEL (four levels) and CONTEXT (two levels) as factors. Because Levene's test for equality of variances was significant for both speakers, using the raw data (Speaker TS: $F(7,94)2.03$, $p<0.05$; Speaker EK: $F(7,94)5.20$, $p<.01$) as well as using

the log-transformed data (Speaker TS: F(7,94)2.54, p<0.05; Speaker EK: F(7,94)4.83, p<.01), variances were assumed to be unequal. Since the interest was not in differences between vowels or in any interaction between vowels and contexts, separate t-tests were run on each of the four vowels to test for the effect of CONTEXT, assuming unequal variances. To counteract the increased chance of finding a significant difference due to fact that we are testing a number of times, the Bonferroni correction was applied to adjust the chances downward. The results for the individual vowels, one-tailed, are given in Table 1. Only in the case of /i:/ for speaker TS was the effect of context not significant.

Table 1. Results of t-tests for the effect of voicing in the post-vocalic consonant on the F1 of four English close vowels, equal variances not assumed, for two speakers. (Speaker TS in the shaded columns and speaker EK, one-tailed, with Bonferroni correction. *=p<.05, **=p<.01.)

	t		df		p	
iː	1.32	3.69	20.3	14.4	n.s.	**
uː	-2.41	2.81	29.0	17.9	*	**
ɪ	2.49	4.59	21.2	19.3	*	**
ʊ	4.97	3.43	17.6	18.5	**	**

The effect of context appears to be the reverse of that predicted by Moreton's *Spread-of-Facilitation Hypothesis*. Not just low vowels are lower before [–voice] codas, but also high vowels. The results for the F1 data at the 25% and 75% time points in the vowels are almost all in the same direction. The difference between the two vowel types is greatest in the middle of the vowel, disconfirming the findings by Van Summers (1987) for open vowels. Like Hillenbrand, Clark, and Neary (2001) and Hawkins and Nguyen's (2004) data, the present data do not lend support to Moreton's hypothesis, since vowels are raised before voiced obstruents regardless of tongue height. Before speculating on alternative explanations, let us now turn to the dialects of Dutch in which similar vowel splits to those reported for English can be observed.

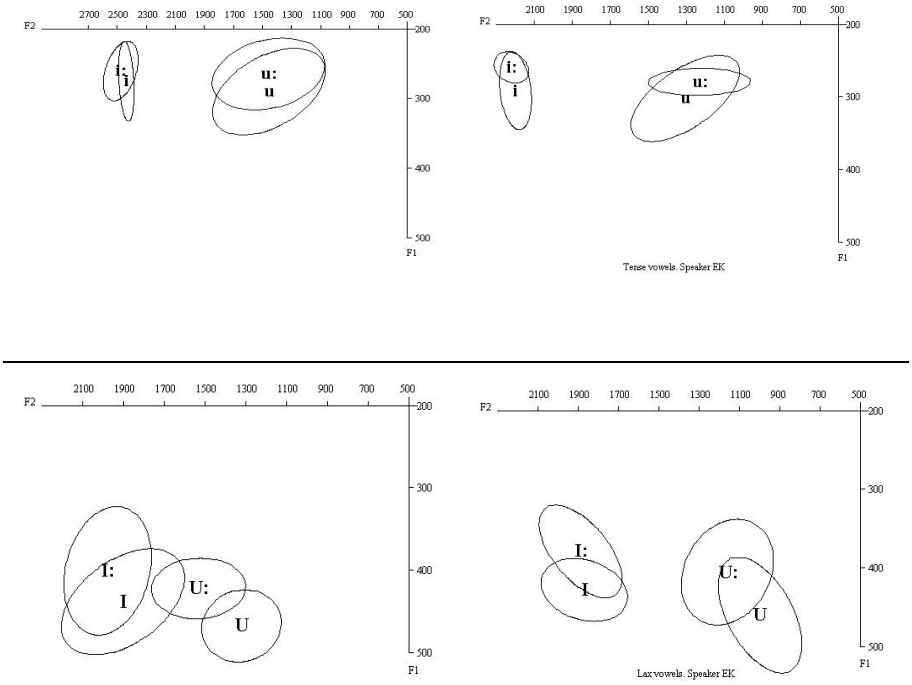


Figure 1. First and second formant plots of front high vowels (panels a and b) and back high vowels (panels c and d) for speakers TP (panels a and c) and EK (panels b and d) separately. V stands for the allophone before [−voice] obstruents and V: for the allophone before [+voiced] obstruents.

3. Vowel quality differences enhancing a Limburgian tone contrast

Vowel quality differences similar to those discussed above have been found in syllables with a tone contrast in Limburgian dialects spoken in the northeast of Belgium and the southeast of the Netherlands. The tone contrast, referred to as Accent 1 vs. Accent 2, occurs in the syllable with word stress and has been described as one between the absence of a lexical H tone (Accent 1) versus the presence of a H (Accent 2) (e.g., Gussenhoven and Aarts 1999). Phonetic realizations vary across the dialects, but frequently reported differences are that syllables with Accent 1 are shorter, have larger F_0 movements, and, less

systematically, have a steeper amplitude decrease towards the end of the sonorant segments in the rhyme. Importantly for our topic, dialects spoken in the southern zone of the Dutch province of Limburg and in Belgium additionally have been reported to have closer vowels in syllables with Accent 2 than in syllables with Accent 1. For instance, in the dialect of Maasmechelen (formerly 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 (2) (Verstegen 1996). Significantly, like German and Dutch, Limburgian has no laryngeal contrast for obstruents in the coda, which are categorically voiceless.

(2) Maasmechelen (Belgian Limburg)

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'

It has likewise been reported that syllables with Accent 1 may be characterized by larger tongue glides than syllables with Accent 2. In the dialect of Maastricht, the diphthongs /ɛi, œy, ɔu/ have markedly different allophones depending on whether they co-occur with Accent 1, as in (3a), or Accent 2, as in (3b) (Gussenhoven and 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 the vowels may variably lose their diphthongal character. The difference is gradient, and native speakers regard the allophones as the same vowel in each of the three cases.

(3) Maastricht (Dutch Limburg)

a. Accent 1: / bɛi/ 'bee', /lœy/ 'people', / dɔuf / 'pigeon': [bɛj, lœj, dɔwf]
b. Accent 2: /bɛi/ 'near', /lœy/ 'lazy', / dɔuf / 'deaf': [bɛːːˢⁱ, lœːːˢʸ, dɔːːˢʷf]

The closer second elements of the diphthongs and the opener realizations of the monophthongs go hand in hand in the development of earlier /iː,yː,uː/ in and around Maastricht. Before /r/, these high long vowels lowered to /eː,øː,oː/ when co-occurring with Accent 1, but remained high when co-occurring with Accent 2. In contexts other than before /r/, they diphthongized when co-occurring with Accent 1, but remained monophthongal when co-occurring with Accent 2, as illustrated in Table 2 (Goossens 1956; de Vaan 2002).

Table 2. Tone-related historical development of high vowels in the dialect of Maastricht (after de Vaan 2002)

	Accent 1			Accent 2		
iː	ɛi	ʃɛif	‘disk’	iː	ɣriːs	‘grey’
yː	øy	bøys	‘tube’	yː	ryːk	‘(he) smells’
uː	ɔu	dɔuf	‘pigeon’	uː	uːt	‘out’
iːr	eː	beːr	‘beer’	iː	ɣiːr	‘stingy’
yːr	øː	døːr	‘dear’	yː	vyːr	‘fire’
uːr	--			uː	zuːr	‘sour’

This chapter makes the important claim that the Limburgian vowel quality adjustments, of which more examples can be given involving both mid and high vowels, resemble the English facts summarized by Moreton (2004). The ‘higher off-glide’ of English pre-[–voice] vowels is to be equated with the larger tongue glides of Limburgian syllables with Accent 1, both favouring a higher endpoint of the diphthong. And the low monophthongs of English can be related to the lower mid and high vowels of Limburgian in syllables with Accent 1. In (4), the formulations are chosen so as to cover both languages.

(4) Accent 1 (Limburg); pre-[–voice] (English): (a) higher off-glide in diphthongs
(b) lower monophthongs
Accent 2 (Limburg); pre-[+voice] (English): (a) lower off-glide (monophthongization)
(b) higher monophthongs

If the way the facts have been collapsed in (4) is correct, an obvious question arises. Why does Accent 1 pair up with the [–voice] rather than the [+voice] coda obstruent? Strikingly, both phonological oppositions are enhanced by duration. Vowels are shorter in syllables with Accent 1 than in syllables with Accent 2, and they are shorter before [–voice] obstruents than before [+voice] obstruents in English.[3] In both English and Limburgian, the durational difference applies not just to the vowel but to the entire sonorant section of the rhyme. It stands to reason, therefore, that if the vowel quality adjustments have a common explanation, this is to be found in the way they may enhance the durational difference. The *Spread-of-Facilitation* hypothesis is thus to be replaced with a *Duration Enhancement Hypothesis,* given in (5).

(5) *Duration Enhancement Hypothesis* (First version):
Monophthongs are raised and diphthongs are monophthongized, because higher monophthongs and less wide diphthongs sound longer than, respectively, lower monophthongs and more diphthongal vowels.

The next section reports on two experiments that sought to establish effects of vowel height and vowel diphthongization on perceived vowel duration. If these effects exist, such that higher vowels and monophthongs sound longer than lower vowels and diphthongs, respectively, hypothesis (5) would clearly be supported.

4. Exploring the Duration Enhancement Hypothesis: Two experiments

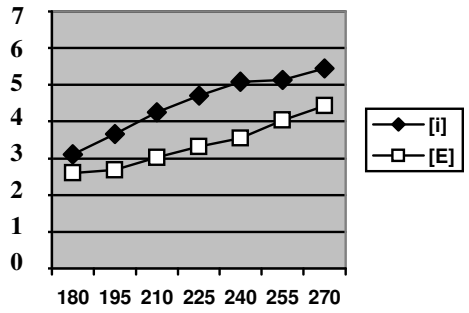
4.1. Experiment 2: The perceived duration of monophthongs and diphthongs

A female speaker of Dutch made digital audiotape recordings of a number of isolated pronunciations of the high vowels [i,y,u], as occurring in Dutch *wie* 'who', *nu* 'now', *koe* 'cow', the mid-open vowels [ɛ,œ,ɔ], as occurring in *bed* 'bed', *oeuvre* 'works', *bot* 'blunt', and the diphthongs [ɛi,œy,ʌu], as occurring in *ei* 'egg', *ui* 'onion', *kou* 'cold' and the vowel [a] as in *na* 'after', pronouncing them with a weakly falling intonation. Unlike Dutch /ɛ,ɔ/, front rounded /œː/ is a long vowel, which was chosen in preference to the more frequent front rounded /ʏ/, because it is opener than this short vowel and thus matches /ɛ,ɔ/ for tongue height. Good tokens of these ten vowels were down-sampled to 16 kHz, monaural files to increase processing speed and save memory space. Using the editing program embedded in Praat (Boersma and Weenink 1992-2002), each of these files was pared down by trimming the edges at zero-crossings or cutting out periods from the central part of the speech file, so as to obtain representative sections of the original speech files that were (close to) 180 ms. in duration. With the help of the option for the manipulation of the fundamental frequency in the Praat package, F_0 was specified to start at 160 Hz. It then rose to 220 Hz in 40 ms., remained at that value for another 40 ms, and then fell for the remainder of the vowel to 110 Hz. Using the option for the manipulation of the duration, each of these ten standardized speech files then served as the basis for six further versions which were 15 ms. apart. This yielded seven durational versions of each vowel:

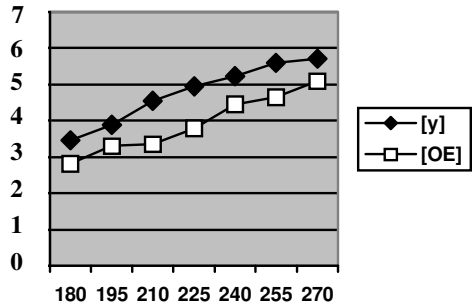
180, 195, 210, 225, 240, 255, and 270 ms. We randomized the 70 stimuli three times and divided these 210 stimuli into blocks of ten. They were up-sampled to 48 kHz and transferred to digital audiotape, making sure that there were no adjacent instances of the same syllable rhyme. In return for a small fee, 34 Dutch listeners, recruited from the student population of the Radboud University Nijmegen, were presented with the stimuli over loudspeakers in a quiet room and asked to rate the duration of each stimulus on a 7-point scale, with the shortest duration appearing on the left of the scale. Because of the difficulty of the task, each stimulus was presented twice with an interval of 700 ms. The interval between one such stimulus presentation and the next was 4.6 s. Each block was preceded by an anchor stimulus of 225 ms. with a schwa-like vowel quality, which corresponded to a scale on the answer sheet in which the box for the fourth scale category had been crossed. Listeners were told that this stimulus represented the mid-point on the scale.

An analysis of variance with DURATIONSTEP (7 levels), VOWELCLASS (three levels: front unrounded, front rounded, back rounded), VOWELHEIGHT (three levels: high, mid, diphthong) showed that all three factors significantly affected perceived duration. Fig. 2 shows the effect of acoustic duration for the monophthongs, but it was consistently present in the diphthongs as well. The data for [a], which were excluded from the statistical analysis, are shown in panel (c), and show the same positive correlation between perceived duration and acoustic duration. Only one of the 42 step increases displayed in Fig. 2 shows a reversed result, with the stimulus for [u] of 270 ms. being heard as shorter than the stimulus of 255 ms. Not surprisingly, these results show that groups of listeners can detect small duration increases, and may serve as a basis of comparison for the effect of other variables. Particularly interesting in the context of our research question is the effect of VOWELHEIGHT. Post-hoc analyses showed that the high vowels are heard as significantly longer than the equivalent mid vowels (i/y/u vs ɛ/œ/ɔ: $F(1,32)$ 18.11, $p<0.01$). However, no overall significant effect was found for the difference between mid monophthongs and the diphthongs (ɛ/œ/ɔ vs ɛi/œy/ɔu: $F(1,32)$ 1.01, ns).[4]

(a)



(b)



(c)

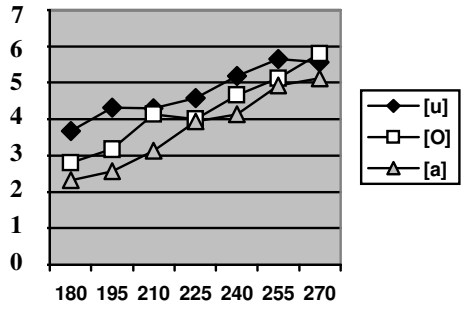


Figure 2. Perceived relative durations by Dutch listeners of seven vowel stimuli with seven acoustic durations, for front unrounded (panel a), front rounded (panel b) and back vowels (panel c) separately. N=34.

4.1.1. Discussion

The results of Experiment 2 confirmed the conjecture that high vowels sound longer than low vowels, but did not confirm the prediction that monophthongs sound longer than diphthongs. If we continue to assume that hypothesis (5) is correct, two questions arise. The first concerns the explanation of the fact that high vowels sound longer than low vowels. The second is why monophthongs failed to show a greater perceived duration than diphthongs.

A conjecture about the answer to the first question was that higher vowels sound longer than lower vowels with the same acoustic duration because hearers compensate for the intrinsically *shorter* duration of high vowels. High vowels tend to be shorter for the physiological reason that the jaw positions required for their production is close to the jaw positions required for the articulation of most consonants (Catford 1977: 197; Maddieson 1997). This explanation will be given a firmer foundation in the general discussion. For now, accepting this explanation to be correct, the second question can be answered by observing that the relation between vowel height and duration has been observed many times in the literature, while there is no generally acknowledged durational relation between diphthongs and long monophthongs (other than that of equality of phonological quantity). There is apparently no natural tendency for diphthongs to be longer than monophthongs, or if there is, it is much smaller than the durational difference between high and low vowels. Dutch, which has a quantity opposition in stressed syllables, is not a representative language for illustrating the relation between vowel height and vowel duration, as high vowels are categorically short. A recent publication on a language with a five-vowel system without quantity contrast (Greek) reports 76 ms. for the close vowels /i,u/, 95 ms. for the mid /e,o/ and 112 ms. for /a/ (Botinis, Fourakis, and Orfanidu 2005: Fig 4.). By contrast, the Dutch diphthongs /ɛi,œy,ʌu/ are 125 ms. in penultimate position in the utterance for the three speakers in Tables 1A, 2A and 3A, as opposed to 122 ms. for the four non-high long vowels /eː,øː,oː,aː/ in Nooteboom (1972), while Rietveld, Kerkhoff, and Gussenhoven (2005) find 166 ms. and 155 ms. for the same classes in word-final utterance-internal position. That is, while the difference between high and low vowels can amount to some 47%, that between monophthongs and diphthongs is less than 5%.

Quite in contrast to the smaller durational difference between monophthongs and diphthongs than that between high and low vowels, the reports in the literature on Limburgian dialects on diphthongization in syllables with Accent 1 and monophthongization in syllables with

Accent 2 are more numerous than those on vowel height differences. This suggests that there is a different, more powerful mechanism underlying the enhancement of vowel shortening by diphthongization. Because the common element in English and Limburgian is not so much diphthongization *per se*, but more specifically the development of a closer endglide, it may be that the shortening effect in the perception is due to the creation of a glide in the position of the second element of the diphthong. By changing, say, [ɛi] into [ɛj] and [ɑu] into [ɑw], the perception of the vowel duration is reduced to [ɛ] and [ɑ] at one go, assuming that the time taken up by the glide is not counted by the listener towards the duration of the vowel.

4.2. Experiment 3: From diphthong to vowel+glide

Experiment 3 intended to compare the pronunciation of diphthongs with vowel-glide combinations, the prediction being that faced with the task of rating vowel duration, listeners will perceive vowel-glide combinations as having shorter vowels, if diphthongs and vowel-glide combinations have identical durations, as reported earlier in Gussenhoven and Driessen (2004). To have a basis for comparison for the perception of vowel+glide combinations, we included [Vm] rhymes, which had the same duration as the diphthongs and vowel+glide combinations. In this experiment we also included high vowels and mid vowels, in an attempt to replicate the earlier finding. Our hypothesis is (7), a more specific version of (5).

(7) *Duration Enhancement Hypothesis* (Final version):
High monophthongs sound longer than low monophthongs because listeners compensate for their inherent shorter duration *(Compensatory Listening)*, and monophthongs sound longer than diphthongs due to an interpretation of the higher off-glide in diphthongs as consonants (*Off-glide Strengthening*).

A female speaker of the dialect of Weert, which contrasts closing diphthongs and phonetically similar vowel+glide combinations (Heijmans and Gussenhoven 1999), recorded three repetitions of the syllable rhymes in Table 3 on digital audiotape. The speaker, who was a first-year language student, was provided with keywords in the dialect's orthography representing the syllable rhymes as well as with their phonetic transcriptions. By exploiting the three-way backness/rounding contrast in the Weert vowel system, we were able to

test our hypotheses three times in the same experiment. In addition, we included the vowel [a], which put the total number of rhymes at 16.

The 3 x 16 (or 48) speech files were down-sampled to 16 kHz and trimmed to 180 ms, as in Experiment 2. In the case of the vowel-glide and vowel-nasal combinations, we took care to obtain approximately equal halves of the 180 ms section for the vowel and the glide or nasal, which seemed adequately to preserve the perceptual difference between diphthongs and vowel+glide combinations. We manipulated the F_0 of all 48 speech files, starting at 160 Hz, rising for 40 ms to 220 Hz, which value was maintained for 40 ms., after which a fall to 110 Hz at the end was created, resulting in a neutral-sounding intonation pattern. Subsequently, the duration of these signals was manipulated so as to produce vowel durations of 160, 200 and 220 ms., in addition to the original 180 ms., for all 48 vowels. The reason why we preferred to use different speech files for the three repetitions of each stimulus rather than identical copies, as in Experiment 2, was that we wanted to minimize the risk of including artefacts in the recordings or subsequent manipulations of the speech files. PSOLA resynthesis of these manipulated files thus yielded 4 (durations) × 3 (repetitions) × 16 (rhymes), or 192 stimuli.

Table 3. Syllable rhymes included in Experiment 3.

V-high	V-mid	VV	VG	Vm
i	ɛ	ɛi	ɛj	ɛm
y	œ	œy	œj	œm
u	ɔ	ɑu	ɑβ	ɑm

We randomized the 192 stimuli, up-sampled them to 48 kHz, and transferred them to digital audiotape, making sure that there were no adjacent instances of the same syllable rhyme. Moreover, we copied five randomly chosen stimuli to appear at the beginning of the test tape and three to appear after the 100th stimulus, where a break was inserted in the test, which now contained 200 stimuli. These were divided into blocks of ten, each preceded by a specially prepared anchor stimulus containing the vowel schwa with a duration 190 ms., the halfway mark between the shortest and longest stimuli.

Each stimulus was presented three times in succession with 700 ms. between repetitions. Each block consisting of the anchor stimulus plus ten stimuli was preceded by a 10 second silent interval followed by a warning signal and 3000 ms. of silence. Twenty-seven judges, all native speakers of Dutch, listened to the tape through loudspeakers in a quiet room and rated the perceived vowel duration of each stimulus on a 7-point scale ranging from ‘very short’ on the left to ‘very long’ on the

right. They were explicitly told that some stimuli consisted of a vowel-consonant combination, in which case they were to estimate the duration of the vowel only. However, they were not told what types of consonant the stimuli might contain.

Scores were averaged per duration step over the three occurrences of each syllable rhyme. An Analysis of Variance was performed on all data except those for [a] with DURATIONSTEP (4 levels), VOWELQUALITY (3 levels), and RHYMETYPE (5 levels, cf. the columns in Table 3). Table 4 gives *F*-ratios, degrees of freedom with Huynh-Feldt corrected degrees of freedom in brackets and 1% significance levels.

Table 4. *F*-ratios, df's and Huynh-Feldt corrected significance levels for DURATIONSTEP, RHYMETYPE and VOWELQUALITY.

	F	df	*p*
DURATIONSTEP	175.94	3 (1.3)	.000
RHYMETYPE	36.20	4 (3.30)	.000
DUR x RHYME	4.20	12 (11.30)	.000
DUR x RHYME x VOW	3.34	24 (24)	.000

The significant three-way interaction DURATION, RHYMETYPE and VOWELQUALITY is due to variation in effect size among the various rhyme types across the three backness/rounding conditions, but other than this unevenness there is considerable consistency in the data. As shown in Fig. 3, acoustic durations correlate consistently with perceived durations. In each of the three vowel quality classes, the high vowel has greater perceived duration than the corresponding mid vowel (cf., the filled and open plot diamonds), and the diphthong has greater perceived duration than the corresponding vowel-glide combination, although in the case of front unrounded vowels the difference is small: the mean scores for [ɛi] and [ɛj] are 3.76 and 3.60, respectively. The vowel-glide combinations may not have sounded quite like a short monophthong followed by a glide in all cases to all listeners, most of whom will have been unfamiliar with the specific vowel-glide rhymes in the experiment, as these are unknown in the standard language. According to Tukey's post-hoc test for homogeneous subsets, a number of individual comparisons proved insignificant at 5%, even though in no case were scores in conflict with our predictions. Table 5 gives the relevant comparisons. The comparisons between [Vm] and all the other rhyme types within each of the three vowel quality classes were significant.

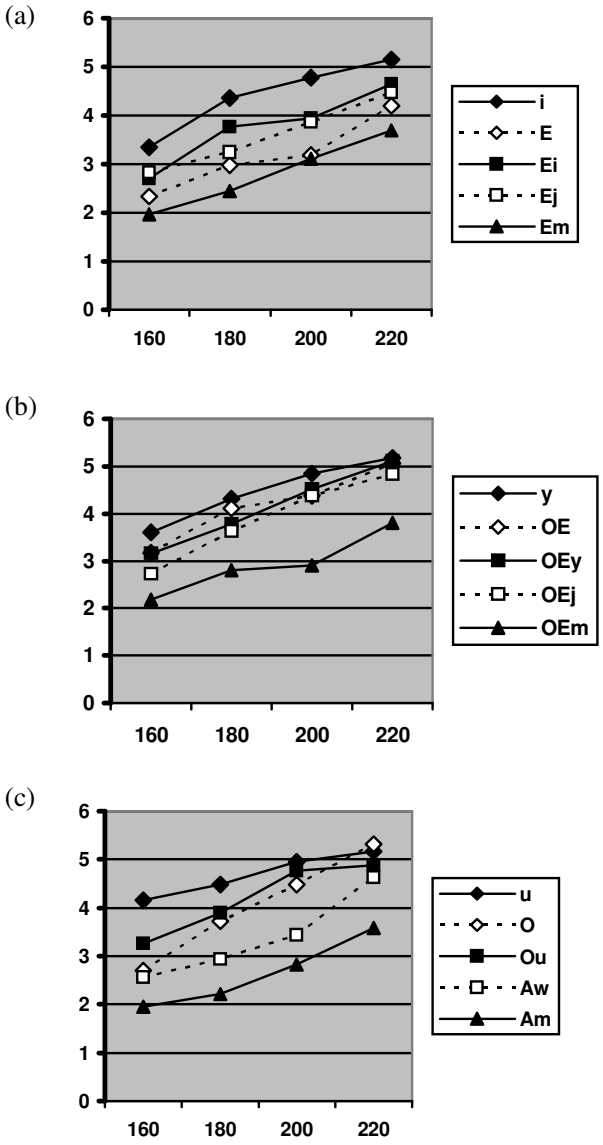


Figure 3. Perceived durations by Dutch hearers of fifteen rhyme vowel stimuli with four acoustic durations, for front unrounded (panel a), front rounded (panel b) and back vowels (panel c) separately. N=27.

Table 5. Relevant results of Tukey's HSD tests of multiple comparisons for the difference in perceived duration between high and low monophthongs and between diphthongs and vowel+glide combinations.

	all vowels	front unround	front round	back round
High vs low monophthongs	*	*	n.s.	*
Diphthong vs vowel+glide	n.s.	n.s.	n.s.	*.

The results of Experiment 3 lend support to the Duration Enhancement Hypothesis (7). First, the effect of vowel height of perceived duration was replicated, and second, we found that there was a consistent trend across three comparisons suggesting that diphthongs have longer perceived vowel durations than phonetically similar vowel+glide combinations, with one of these comparisons reaching significance.

5. Conclusion

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 (Catford 1977). 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 are therefore inherently longer than high vowels. When assessing the duration of a vowel, he will subtract this inherent portion in the duration, before constructing the perceived duration. Putting it differently, unlike high vowels, low vowels include a component in their *articulatory duration* which is obtained as an unintended by-product of the articulatory lowering gesture of the tongue and jaw. By way of compensation, the hearer reduces the acoustic duration when estimating the perceived duration.

This 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 linemarking 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 compensates for the declination in production, bringing a late peak back up to the level it would have had if there had been no declination. A second case is Silverman (1987), who demonstrated with British English listeners that the prominence of the same fundamental frequency peak was lower when combined with the vowel /iː/ than when combined with the vowel /ɑː/. He did this by having hearers select the word with the greatest prominence in utterances with two prominent words, one containing /iː/ and the other containing /ɑː/. The stimuli consisted of pairs of utterances like *They only FAST before FEASTing* and *They only FEAST before FASTing,* which had approximately the same overall level of prominence and in which the two vowels had been cross-spliced. In the case of /iː/, the cross-over point between 'first word most prominent' and 'second word most prominent' occurred if the second peak was 1.7 Hz higher than the first, while if the second peak occurred in a syllable containing /ɑː/, the cross-over point was -6.7 Hz, a difference of 8.3 Hz.

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, listeners don't expect F_0 to be very high in an accented word occurring late in the utterance, because they know that F_0 becomes increasingly lower as the utterances progresses.. In the second case, listeners know that high vowels have higher F_0, for which the most plausible explanation is that articulation of the high vowel causes an upward and, in the case of front vowels, forward pull of the tongue root on the thyroid, causing some tensing of the vocal folds, which as a result vibrate a little faster (cf., Silverman 1987: ch 3; Maddieson 1997). The effect reported in this chapter adds to these cases, in that yet another articulatory advantage, the longer duration of open vowels, is subtracted from the acoustic value before a perceptual judgement is made.

The difference between monophthongs and diphthongs was likewise argued to arise from a desire to enhance the perception of a duration difference. But while the *motivation* for the strengthening of the second element in the short allophones is the same as that for lowering the vowel, i.e., to enhance the durational contrast, the effect is in no way dependent on compensatory listening. Rather, by strengthening the off-glide, the second element of the diphthong is perceived as a glide, and as such is not included in the hearer's percept of the vowel. As a result of this *Off-glide Strengthening*, the perceived vowel duration is reduced. This then is the explanation for the fact that the allophonic difference between the second elements of closing diphthongs before

[−voice] and [+voice] obstruents is greater than that between the first elements. (Recall that in Experiment 1 the greatest difference between the allophones of monophthongs was found in the middle of the vowel, not at the end.)

The explanation offered by Moreton (2004) of the lower low vowels and the higher diphthongal endglides before [−voice] as an effect of the leakage of hyperarticulation of the −voice] consonant to the last part of the preceding monophthong or diphthong would, at first sight, appear appealing because of its unifying nature: the peripheralization of vocalic articulations. Also, as pointed out by John Kingston, hyperarticulation as observed in accented syllables in comparison with unaccented syllables does indeed have effects that are consistent with Moreton's assumption. Mouth openings are wider for all vowels, and low vowels are lower, but at the same time tongue positions for high front vowels are more front or higher or both (Harrington, Fletcher and Beckman 2000; Erickson 2002). By contrast, the present explanation needs to call on two mechanisms to account for the same facts. Nevertheless, the hypothesis that vowel lowering and off-glide strengthening are two different ways of making vowels sound shorter has a number of advantages over Moreton's *Spread-of-Facilitation Hypothesis*.

1. Unlike the *Spread-of-Facilitation Hypothesis,* the *Duration Enhancement Hypothesis* is capable of explaining the vowel quality adjustments both as an enhancement of the English laryngeal contrast and as an enhancement of the Limburgian Dutch tone contrast, on the grounds that both phonological contrasts are enhanced by duration. Significantly, two languages appear to employ the compensatory perception effect to enhance phonological contrasts whose main phonetic exponent is a durational difference, even though the *phonological* contrasts concerned are quite different. If this explanation is correct, it is to be expected that, in general, phonological duration contrasts show the same correlation between duration and vowel quality. Labov's treatment of chain shifts involves just these regularities: long vowels raise (Principle I), a correlation that has been observed in numerous languages (Labov 1994: 122) and short vowels lower (Principle II) (Labov 1994: 116). Although Labov (2001) offers no explanation for these correlations, his suggestion that these effects are due to contrast enhancement is in keeping with the explanation offered here.

2. The *Spread-of-Facilitation Hypothesis* incorrectly predicts that high vowels are higher before [-voice] obstruents. By contrast, the *Duration Enhancement Hypothesis*, which makes the opposite prediction, readily finds support in independent data, as well as in the

results of Experiment 1. The results of Experiments 2 and 3 suggest that vowel raising and off-glide strengthening reduce perceived vowel duration, even though the results for off-glide strengthening reached significance in only one of the three comparisons. Additional circumstantial evidence can be found in the segmental phonologies of Limburgian dialects which contrast vowel+glide combinations with phonetically similar monophthongs and diphthongs. Thus, the dialect of Weert has contrasts like /βæjc/ '(the wind) blows', /lɛit/ 'sorrow', /dœjts/ 'German' (adj.), /kœyt/ 'fun', /ɑβx/ 'eye', /ʌux/ 'also' (Heijmans and Gussenhoven 1998). Unlike what is often assumed, therefore, it is not the case that an analysis of some diphthong [ai] as either /ai/ or /aj/ is immaterial. Representationally, some provision will have to be made for the dialect of Weert either for a featural difference between glides and vowels or for the inclusion of a coda constituent, since a bare moraic representation as in Hayes (1989) would otherwise not express the contrast.

3. 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. Moreton addresses this point, making it clear that the articulation of fortis consonants involves greater effort than that of lenis consonants. However, there are no indications that the lenis articulation, which involves quite considerable lengthening of preceding sonorant portions in addition to further measures like velic leakage and cavity expansion to maintain the transglottal pressure, requires less articulatory control.

4. Instead of a disadvantage, the fact that two mechanisms have been found to lie at the basis of a common goal may be seen as the expected situation. Contrast enhancement is best served by the recruitment of different mechanisms aspiring to achieve the same effect, that of making the contrast more salient. Preglottalization of [−voice] obstruents is a different mechanism than shortening the sonorant portion of the rhyme, but does serve the same goal of making the sonorant portion of the rhyme sound short. Hawkins and Nguyen (2004) found a difference in the pronunciation of onset /l/ between syllables closed by [−voice] obstruents, where onset /l/ is clearer, and [+voice] obstruents, where it is darker. This can likewise be interpreted as yet another attempt by speakers to signal the shorter pre-[−voice] duration, since, as a coda consonant, dark [l] is longer than the clear [l] of the onset. Enhancement thus appears to be a collusive enterprise in which speakers insert hints in their pronunciation of words with precarious contrasts that are based on phonetic correlations between the

primary feature and the enhancing articulatory parameter (here, preceding segment duration), and between the enhancing articulatory parameter and further phonetic parameters.

Finally, there is the hairy issue to what extent the explanation for long vowel raising, short vowel raising, pre-fortis clipping, preglottalization, etc. are teleological. What is clear is that enhancement features are non-random, in the sense that they aid rather than mask the contrast at issue. Also, if the idea of 'phonetic knowledge' is accepted (Kingston and Diehl 1994), they must be introduced for a purpose, even though such knowledge is tacit, like most if not all linguistic knowledge. The alternative assumption that random variation will at times gives rise to favorable effects in the speech production process does not stave off the conclusion that speech behavior is purposeful, since for speakers to be able to identify the process at issue and to adopt it in subsequent pronunciations inevitably implies they do so for a reason. However, this short-term teleology must not be taken to imply that speakers strive towards typologically unmarked grammars. It only means that they can speak clearly if they need to. What this chapter has intended to show is that the ways in which they do this may be more ingenious than previously thought.

Acknowledgements

I would like to thank Femke Dekkers, Wilske Driessen, Joop Kerkhoff, and Marco van de Ven for their hard work on the experiments reported or referred to here, without whose competence and perseverance this research would not have been done. As ever, I am greatly indebted to Toni Rietveld for his statistical help, which was equally indispensable for the publication of the results. I thank Flor Aarts for his patience with my continued requests for information about the facts of the Maastricht dialect. And if this contribution strikes the reader as tolerably digestible and suitably placed within a wider framework of efforts to understand the details of the speech process, then this is in no small measure due to the comments by John Kingston and the editors of the volume on an earlier version.

Notes

1. As John Kingston points out, not all enhancement requires that the speaker inferences the role of the added features on the basis of phonetic knowledge about articulatory correlations, since enhancement may directly boost the effect of the primary feature, such as when lips are rounded to enhance the acoustic effect of tongue backing, or when F_0 and

F1 are lowered around the edges of voiced obstruents to suggest the presence of voicing (Kingston & Diehl 1995).

2. The lower F1 in vowels after voiced consonants than after voiceless consonants is an independent effect. I have no explanation for this finding.
3. Extensive motivation for the analysis of the Limburgian word prosodic contrast as a phonological tone contrast enhanced by duration, rather than a quantity contrast that is enhanced by F_0, is presented for the dialect of Cologne by Gussenhoven & Peters (2004).
4. Separate paired comparisons between the low monopthongs and the diphthongs did reveal that the difference between [ɔ] and [ɔu] was significant (F(1,32) 14.051, p<0.01). Also, the interaction between Diphthong and VOWELCLASS was significant [F(1.76,64) 23.95, p<0.01]. I refrain from attempts to explain these results.

References

Boersma, Paul and Dirk Weenink
1992-2002 *Praat: Doing Phonetics by Computer.* <www.praat.org>

Botinis, Antonis, Marios Fourakis, and Joanna Orfanidou
2005 Vowel durations of normal and pathological speech. Proceedings FONETIK 2005. Gothenburg: Göteborg University, Department of Linguistics. 123-126.

Catford, J.C.
1977 *Fundamental Problems in Phonetics*. Edinburgh: Edinburgh University Press.

de Vaan, Michiel
2002 Wgm *i and *u voor r in Zuid-Limburg. *Taal en Tongval* 54: 171-182.

Erickson, Donna
2002 Articulation of extreme formant patterns for emphasized vowels. *Phonetica* 59: 134-149.

Goossens, Jan
1956 Stoottoon en diftongering van Wgm. î and û in Limburg. *Taal en Tongval* 8: 99-112.

Gussenhoven, Carlos and Flor Aarts
1999 The dialect of Maastricht. *Journal of the International Phonetic Association* 2: 55-66.

Gussenhoven, Carlos and Wilske Driessen
2004 Explaining two correlations between vowel quality and tone: The duration connection. In *Prosody 2004,* B. Bel & I. Marlien (eds.), 179-182.

Gussenhoven, C. and Jörg Peters
2004 A tonal analysis of Cologne *Schärfung*. *Phonology* 21: 252-285.

Harrington, Jonathan, Jane Fletcher, and Mary E. Beckman
2000 Manner and place conflicts in the articulation of accent. In *Papers in Laboratory Phonology V: Acquisition and the Lexicon*, M. Broe & J.B. Pierrehumbert (eds.), 40-51. Cambridge: Cambridge University Press.
Hawkins, Sarah and Nöel Nguyen
2004 Influence of syllable-coda voicing on the acoustic properties of syllable-onset /l/. *Journal of Phonetics* 32: 199-231.
Hayes, Bruce
1989 Compensatory lenghtening in moraic phonology. *Linguistic Inquiry* 20: 253-306.
Heijmans, Linda and Carlos Gussenhoven
1998 The Dutch dialect of Weert. *Journal of the International Association* 28: 107-112.
Heijmans, Linda
2003 The relationship between tone and vowel length in two neighboring Dutch Limburgian dialects. In *Development in Prosodic Systems*, P. Fikkert & H. Jacobs (Eds.), 7-45. New York: Mouton de Gruyter.
Hillenbrand, James M., Michael J. Clark, and Terrance M. Nearey
2001 Effects of consonant envioronment on vowel formant patterns. *Journal of the Acoustical Society of America* 109:, 748-763.
Kingston, John and Randy L. Diehl
1994 Phonetic knowledge. *Language* 70: 419-454.
1995 Intermediate properties in the perception of distinctive feature values. In *Phonology and Phonetic Evidence: Papers in Laboratory Phonology IV*, B. Connell & A. Arvaniti(eds.), 7-27. Cambridge: Cambridge University Press.
Labov, William
1994 *Principles of Linguistic Change. Volume 1: Internal Factors*. Malden, MA and Oxford, UK: Blackwell.
2001 *Principles of Linguistic Change. Volume 2: Social Factors*. Malden , MA and Oxford, UK: Blackwell.
Maddieson, Ian
1997 Phonetic universals. In *The Handbook of Phonetic Sciences,* W.J. Hardcastle & J. Laver (eds.), 619-639. Oxford: Blackwell.
Moreton, Elliot
2004 Realization of English postvocalic [voice] contrast in F1 and F2. *Journal of Phonetics* 32: 1-33.
Nooteboom, Sieb G.
1972 *Production and Perception of Vowel Duration: A Study of Durational Properties of Vowels in Dutch.* PhD dissertation, Rijksuniversiteit Utrecht.

Rietveld, Toni, Joop Kerkhoff, and Carlos Gussenhoven
2004 Word-prosodic structure and vowel duration in Dutch. *Journal of Phonetics* 32: 349-371.
Silverman, Kim
1987 *The Structure and Processing of Fundamental frequency Contours*. PhD Dissertation. University of Cambridge.
Stevens, Kenneth N. and Samual J. Keyser
1989 Primary features and their enhancement in consonants. *Language* 65: 81-106.
Thomas, Erik R.
2000 Spectral differences in /ai/ offsets conditioned by voicing of the following consonant. *Journal of Phonetics* 28: 1-25.
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, John C.
1981 *Accents of English*. Three volumes. London: Longman.
Wolf, C.G.
1978 Voicing cues in English final stops. *Journal of Phonetic 6*: 299-309.