ISCA Archive
http://www.isca-speech.org/archive

Speech Prosody 2004
Nara, Japan
March 23-26, 2004

Explaining Two Correlations between Vowel Quality and Tone: The Duration Connection

Carlos Gussenhoven & Wilske Driessen

Centre for Language Studies
University of Nijmegen, The Netherlands
c.gussenhoven@let.kun.nl

Abstract

High vowels sound longer than low vowels when acoustic durations are equal. Also, the perceived vowel duration of diphthongs is longer than that of phonetically similar vowel-glide combinations. These experimental findings lie at the basis of two vowel splits that took place in tonal dialects in the Netherlands and Belgium in which duration is used as enhancement of a tone contrast.

1. Introduction

A number of dialects in the Dutch and Belgian provinces of Limburg have developed vowel quality distinctions by way of enhancement or by way of a reinterpretations of a tone contrast. In the more conservative tonal dialects, a distinction between 'Accent 1' and 'Accent 2' exists, which occurs on segmentally identical rhymes of stressed syllables and which can be illustrated with the minimal pairs in (1) for the dialect of Maastricht [1]. In isolation, Accent 1 is a fall from high to low in this dialect, while Accent 2 is either a mid level or a weak rise to mid, in both cases followed by a late fall.

(1)	taːk^1	'task'	taːk^2	'roof'
	bɑl^1	'ball, dance'	bɑl^2	'ball, toy'
	peeʀt^1	'horse-PL'	peeʀt^2	'horse-SG'

There are two ways in which the tone contrast has given rise to vowel quality differences.

- Diphthongal vowels split into a more diphthongal vowel, which in its extreme form is a vowel-glide combination, in syllables with Accent 1 and a less diphthongal vowel in syllables with Accent 2.
- Mid monophthongs split into a lower vowel in syllables with Accent 1 and a higher vowel in syllables with Accent 2.

For instance, the dialect of Weert has split the diphthongs /ɛi, œy, ɔu/ into /ɛi, œy, ɔu/ and /æj, œj, ɑβ̞/, where the vowel-plus-glide combinations represent a more extensive tongue glide and correspond etymologically with Accent 1 [2]. In (2) some examples are given.

(2)	stæjn	'stone-PL'	stɛin	'stone-SG'
	bœjm	'tree-PL'	bœym	'tree-SG'
	ɑβ̞x	'eye'	ʌux	'also'

The 'diphthongal' connection between tone and quality arose not just through a reinterpretation of the tone contrast, but also through differential sound changes involving diphthongization of long vowels. For instance, in the dialects spoken in a string of villages in the Haspengouw area of the Belgian province of Limburg, long high vowels are undergoing a process of diphthongization which is generally more advanced in syllables with Accent 1 than in syllables with Accent 2 [3]. Likewise, in the tonal dialect of Sittard in the Netherlands, a similar change was completed in syllables with Accent 1, while syllables with Accent 2 still have the original monophthong [4]. This is shown in (3).

(3)	kɛin^1	'chink'	kee^{2}nə	'to bud'
	ʃmɛi^1	'smith-PL	ʃmeet^2	'smith-SG
	blɔut^1	'blood'	noot^2	'nut'
	hɔun^1	'hen'	zoon^2	'son'
	bœy^{1}kəʀ	'book-PL'	køø^{2}kə	'kitchen'
	vœy^{1}lə	'to feel'	vøø^{2}lə	'foal, filly'

The second connection is rarer. Verstegen presents data for his native dialect of Mechelen-aan-de-Maas [5]. A number of tonally distinguished word forms with mid vowels have split into an opener and closer type, as illustrated in (4).

(4)	ɣɛɛl^1	'yellow-ATTR'	ɣeel^2	'yellow-PRED'
	wɛɛx^1	'road-PL'	weex^2	'road-SG'
	ɣɔɔn^1	'go-1SG,PRES'	ɣoon^2	'go-1PL,PRES'
	nɔɔl^1	'needle-SG'	noo^{2}lə	'needle-PL'

In other cases, the tone contrast was given up after the split. Cajot [6] shows that Belgian dialects that have lost the tonal contrast tend to have larger numbers of tone-induced vowel splits than dialects in which the tone contrast still exists.

1.1. Towards an explanation

There has been an attempt at a psycho-acoustic explanation of the fact that in contrast to Accent 2 Accent 1 is characterized by a (greater) vowel quality glide and a greater pitch glide. [7] hypothesized that tone and resonance glides are processed differently depending on their duration. The results of their experiment were negative. There was however an important element in Peeters-Schouten attempt to understand the correlation between diphthongization and Accent 1. They assumed that it might be mediated by the phonetic duration difference that exists between the two words tones. That is, the correlations between tone and vowel quality may be indirect. Indirect correlations between tone and vowel quality are known to arise through voice quality differences. Breathy voice is associated with lower F_0 as well as with ATR in Mon-Khmer languages, while for tense voice the opposite relations hold [8]. Since [+ATR] may lead to closer vowels than [-ATR], a correlation between close vowels and low tone may arise. A correlation of

10.21437/SpeechProsody.2004-42

this type is presented by Khmer and Vietnamese, where Khmer reinterpreted a voice quality contrast as a vowel height contrast, but Vietnamese as a tone contrast.

Since there are no typologically convincing correlations between level tone (Accent 2) and monophthongization or between level tone (Accent 2) and vowel raising, an opportune question is whether there might be a third, mediating factor in the relation between tone and vowel quality in the Limburg dialects, just as voice quality is a mediating factor in Mon-Khmer languages.

The connection between tone and duration in Limburg tone has been widely reported, and goes back to the tonogenesis, according to [9]. It holds that, around 1300, speakers of the Cologne dialect were caught between the sociolinguistic pressure to lengthen the vowels of monosyllabic singular stems and the need to keep the resulting lengthened vowels distinct from the phonologically long vowels in otherwise identical plural forms. The need to preserve the morphological distinction led to a drawled pronunciation in the singular, which was distinct from the fall used for the plural, assuming citation pronunciations. The resulting analysis as HHL for the singular versus HL for the plural led to the introduction of a lexical H-tone in singular forms, i.e. Accent 2 (where HL was regarded as the intonational pitch accent). It is against this background of simulating long vowels that we need to understand the correlations between vowel quality and tone:

- Phonetic and phonological features involved in the realization of the contrast between Accent 1 and 2 are due to enhancement of a duration contrast.

1.2. The hypotheses

A series of hypotheses and perception experiments conducted with Dutch listeners which tested these hypotheses eventually led to the formulation of two hypotheses which we believe correctly explain the two correlations between Accent 1 on the one hand and vowel lowering and diphthongization on the other.

1.2.1. The relation between diphthongization and duration

By changing a diphthong into a vowel-glide combination, the perceived vowel duration is dramatically shortened as a result of the fact that the second element of the diphthong is no longer included in the vowel percept, since it now represents a coda consonant. That is, in their desire to create shorter vowel durations for diphthongs in syllables with Accent 1, these diphthongs were restructured to vowel-glide combinations. This change is responsible for the widespread occurrence in these dialects of contrasts between diphthongs and vowel-glide combinations that are phonetically similar, as in Weert. This is a typologically highly exceptional type of contrast.

Thus, originally, in order to make diphthongs sound short, speakers transformed them into short monophthongs, and gave the second element of the diphthong a consonantal status, i.e. [ɛi] → [ɛj]. The monophthongization of diphthongs with Accent 2 is then to be explained as a way of enhancing the contrast with Accent 1. Later, the correlation became an arbitrary element in the complex of features that realize the tone contrast.

1.2.2. The relation between vowel height and duration

The hypothesis that captures the correlation between vowel height and duration is very different, and based on the assumption that listeners compensate for phonetic parameters which are naturally activated by specific articulatory actions. For instance, raised tongue height leads to an upward pull at the larynx, causing stiffer vocal folds and higher F_0. Listeners have been shown to ignore the effects of this intrinsic F_0 of vowels by [10]. We explain this phenomenon of 'compensatory listening' as the listener's subtraction from an unavoidably obtained boost of a phonetic parameter, F_0 in Silverman's example, caused by an articulatory action, in this case the production of a high vowel. Similarly, the declination effect first demonstrated for English by [11] can be explained as the subtraction of a boosted F_0 at the beginning of a breath group, when subglottal pressure and F_0 are higher than average. As a result of the subtraction, earlier peaks have lower pitch than later peaks. Figure 1 shows this line of thinking in a graphic form.

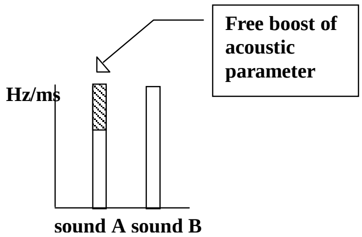

Figure 1: *Compensatory listening as the listener's subtraction of an increase of an acoustic parameter obtained by the speaker as an unintended by-effect of some articulatory action. The unintentional boost of the acoustic parameter in sound A therefore leads to a reduced value in the perception of the corresponding auditory correlate as compared to sound B, which has an identical value for the acoustic parameter concerned.*

If we follow the same reasoning in the case of vowel height and duration, higher vowels must sound longer than lower vowels. Since lower vowels are intrinsically longer due to the larger distance the tongue body needs to cover from the consonantal position to the articulatory vowel target, the listener will deduct some of the extra travelling time from the acoustic measure before perceiving the phonetic duration. Speakers of the tonal dialects must have exploited this fact in their attempts to make vowels with Accent 2 sound longer, by making them higher than vowels with Accent 1.

These explanations lead to two hypotheses:

1. The perceived duration of diphthongs is greater than that of vowel-glide combinations if acoustic durations are identical.
2. The perceived duration of high vowels is greater than that of mid vowels if acoustic durations are identical.

2. An experiment

In order to test these hypotheses, we needed to collect perceived vowel duration judgements for stimuli that varied along the relevant dimensions (a) diphthong vs vowel-glide, and (b) tongue height. Because we expected vowel-glide stimuli to be interpreted as vowel-plus-coda structures, we also included stimuli that consisted of a vowel and an unambiguous consonant, by way of baseline. These stimuli should have identical acoustic durations, F_0 contours and comparable intensity levels, so that only the spectral information varied.

2.1. Method

A female speaker of the dialect of Weert recorded three repetitions of the syllable rhymes in Table (1) using a Sony DAT Deck (DTC-ZA5ES; stereo aifc-files, sampling frequency 48kHz). 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.

Table 1: *Front unrounded, front rounded and back rounded syllable rhymes consisting of (a) high vowels, (b) low vowels, (c) closing diphthongs, (d) vowel-glide combinations and (e) vowel-nasal combinations.*

V-hi	V-lo	VV	VG	VC
i	ɛ	ɛi	ɛj	ɛm
y	œ	œy	œj	œm
u	ɔ	ɑu	ɑβ̞	ɑm

The motivation for having the speaker record each syllable rhyme three times rather than obtaining three copies of the same recording was that we wanted to minimize the risk of including artifacts in the pronunciation or the subsequent manipulation of the speech files. Since our hypotheses concern potentially subtle effects, such care seemed to us to be called for.

2.2. Stimulus preparation

The (3 $\times$ 16) 48 speech files were down-sampled to 16kHz, monaural files to increase processing speed and save memory space. Using the editing program embedded in Praat [12], each of these files was pared down by cutting off the edges at zero-crossings so as to obtain representative sections of the original speech files that were the nearest to 180 ms in length. In no case did we need to lengthen the speech file. Moreover, 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. We manipulated the F_0 of all 48 speech files, choosing 160 Hz at 0 ms, 220 Hz from 40 to 80 ms, and 110 Hz at 180 ms, which resulted in a neutral (falling) intonation pattern. We subsequently manipulated the duration of these signals with the help of the same menu option 'Manipulation', so as to produce 48 versions of 160 ms, 200 ms and 240 ms each. By ordering the F_0 and duration manipulations in this way, we compressed and stretched the F_0 contour along with the waveform, thus avoiding decisions as to how the F_0 contour was to be kept constant across stimuli of different durations. While we were aware that the JND for vowel duration will lie around 20 ms, the choice of a 20 ms step size was motivated by the expectation that the effect we are looking for may be small and that high accuracy in the establishment of perceived duration was essential. A PSOLA resynthesis of these manipulated files thus yielded 4 (durations) $\times$ 3 (repetitions) $\times$ 16 (vowels), or 192 stimuli.

2.3. The test

We randomized the 192 stimuli, up-sampled them to 48 kHz, and transferred them to a Digital Audio Tape, 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 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, and each block was 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. The ISI was 4500 ms. Each block of 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 took part in the experiment on the basis of a small fee. They were recruited from the student population of the University of Nijmegen, 19 females and eight males, in the age range of 19 to 29 years. In an instruction sheet, they were asked to rate the duration of the vowel in each stimulus on a 7-point scale ranging from 'very short' on the left to 'very long' on the right, and to use the anchor stimulus as their reference point. Each occurrence of the anchor stimulus corresponded with a scale on which the fourth point, the halfway mark, was checked off. They listened to the DAT tape through headphones in a language laboratory. The experiment was divided into two parts of 15 minutes each, so that judges could have a break.

3. Results

Scores were averaged per duration step over the three occurrences of each syllable rhyme for all syllable rhymes except [a]. An Analysis of Variance was performed on the data with DURATIONSTEP (4 levels), RHYMETYPE (5 levels, cf. the columns in Table 1) with VOWELQUALITY (3 levels) nested under RHYMETYPE. Table 2 gives F-ratio's, df's with Huynh-Feldt corrected df's in brackets of effects with 1% significance levels.

Table 2: *F-ratio's, df's and Huynh-Feldt corrected significance levels for DURATIONSTEP, RHYMETYPE and VOWELQUALITY*

	F	df	$p \leq 0.001$
DURATIONSTEP	175.94	3 (1.3)	.000
RHYMETYPE	36.20	4 (3.30)	.000
DUR $\times$ RHYME	4.20	12 (11.30)	.000
DUR $\times$ RHYME $\times$ VOW	3.34	24 (24)	.000

The significant three-way interaction with VOWELQUALITY is resolved in Table 3. The interactions appear to be due to variation in effect size among the various rhyme types across the three backness/rounding conditions, but there is considerable consistency in the data. As shown in Figure 2, in all three cases, the high vowels have greater perceived duration than the mid vowels (cf. the open plot characters), and the vowel-glide combinations have shorter durations than the diphthongs, while being longer than the vowel-nasal rhymes.

4. Discussion

The results of our experiment are entirely in line with the hypotheses we formulated earlier. Higher vowels sound longer than lower vowels and the vowel duration of diphthongs is greater than that of vowel-glide combinations. We suggest that these facts explain the vowel splits that took place in the tonal dialects of the Belgium and the Netherlands.

Table 3: *F-ratio's, df's and Huynh-Feldt corrected significance levels for DURATIONSTEP and RHYMETYPE, separately for VOWELQUALITY*

Front Unrounded	F	df	p≤ 0.001
DURATIONSTEP	104.32	3 (1.83)	.000
RHYMETYPE	26.07	4 (3.93)	.000
Front Rounded			
DURATIONSTEP	123.51	3 (2.06)	.000
RHYMETYPE	29.59	3 (2.06)	.000
Back Rounded			
DURATIONSTEP	102.13	3 (1.89)	.000
RHYMETYPE	25.86	4 (3.34)	.000
DUR × RHYME	7.20	12 (12)	.000

5. References

[1] Gussenhoven, C.; Aarts, F., 1999, The dialect of Maastricht *JIPA*, 29, 155-165.

[2] Heijmans, L., 2003, The relationship between tone and vowel length in two neighbouring Dutch Limburgian dialects. In *Development in Prosodic Systems*, P. Fikkert; H. Jacobs (eds). Berlin: Mouton de Gruyter, 7-45.

[3] Peeters, W.; Schouten, M., 1989, Die Diphthongierung der westgermanischen î- und û-Laute im Limburgischen, *Zeitschrift für Dialektologie und Linguistik*, LVI, 309-318.

[4] Dols, W., 1953, *Sittardse diftongering: een hoofdstuk uit de historische grammatica*. Sittard: Alberts Drukkerijen.

[5] Verstegen, V., 1996, Bijdrage tot de tonologie van Oost-limburgse dialecten. In *Een eeuw Limburgse dialectologie*, H. van de Wijngaard (ed), Hasselt/Maasticht: VLDN/Vereniging Veldeke Limburg, 229-234.

[6] Cajot, J., 2001, Een toonloze enclave in polytoon gewaand gebied. Een les in structurele fonologie?, *Jaarboek van de Vereniging voor Limburgse Dialect- en Naamkunde*, 3, 71-88.

[7] Schouten, M.E.H.; Peeters, W.J.M., 2000, Searching for an explanation for diphthong perception: Dynamic tones and dynamic spectral profiles, *Phonetica*, 57, 17-39.

[8] Yip, M., 2002, *Tone*. Cambridge: Cambridge Universty Press.

[9] Gussenhoven, C., 2000, On the origin and development of the Central Franconian tone contrast. In *Analogy, Levelling, Markedness: Prinicples of Change in Phonology and Morphology*, A. Lahiri (ed.), 213-260.

[10] Silverman, K.E.A., 1987, *The structure and processing of fundamental frequency contours.* PhD Thesis, Cambridge University.

[11] Pierrehumbert, J.B., 1979, The perception of fundamental frequency declination, *JASA*, 66, 363-369

[12] Boersma, P.; Weenink, D., 1992, *Praat. Doing Phonetics by Computer*, Version 4.0. www.praat.org

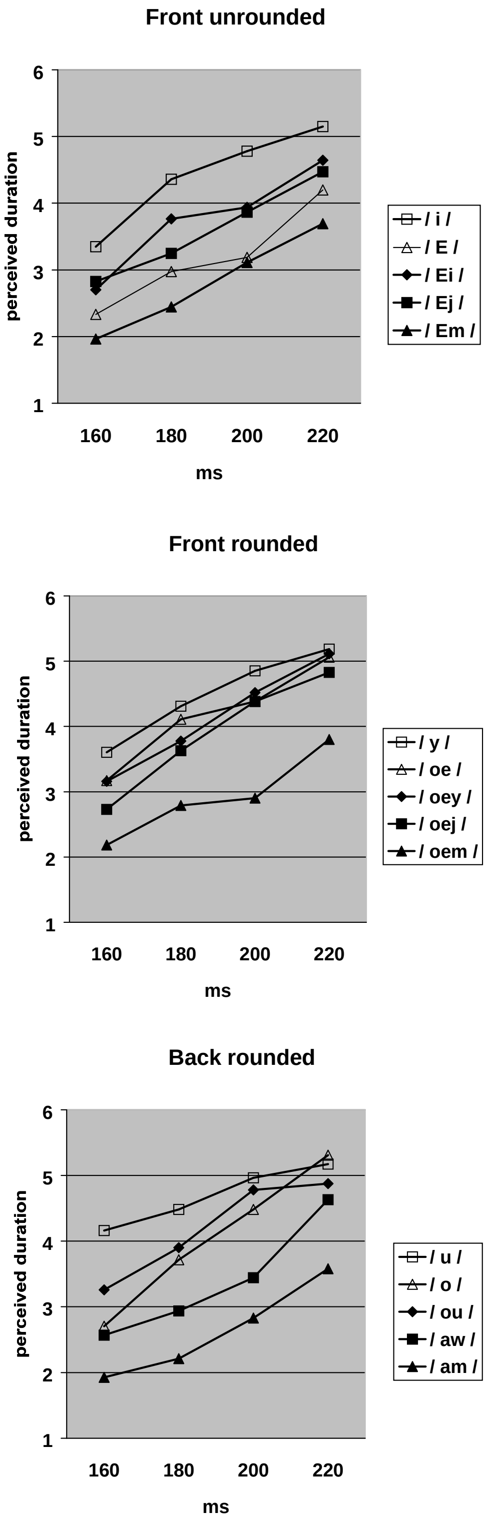

Figure 2: *Perceived vowel duration as a function of acoustic duration and rhyme type for front unrounded, front rounded and back rounded vowels separately.*