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

Speech Prosody 2002
Aix-en-Provence, France
April 11-13, 2002

# Language-specific Uses of the Effort Code

*Aoju Chen, Carlos Gussenhoven and Toni Rietveld*

Centre for Language Studies
University of Nijmegen, The Netherlands
aoju.chen@let.kun.nl;c.gussenhoven@let.kun.nl;a.rietveld@let.kun.nl

## Abstract

Two groups of listeners with Dutch and British English language backgrounds judged Dutch and British English utterances, respectively, which varied in the intonation contour on the scales EMPHATIC vs. NOT EMPHATIC and SURPRISED vs. NOT SURPRISED, two meanings derived from the Effort Code. The stimuli, which differed in sentence mode but were otherwise lexically equivalent, were varied in peak height, peak alignment, end pitch, and overall register. In both languages, there are positive correlations between peak height and degree of emphasis, between peak height and degree of surprise, between peak alignment and degree of surprise, and between pitch register and degree of surprise. However, in all these cases, Dutch stimuli lead to larger perceived meaning differences than the British English stimuli. This difference in the extent to which increased pitch height triggers increases in perceived emphasis and surprise is argued to be due to the difference in the standard pitch ranges between Dutch and British English. In addition, we found a positive correlation between pitch register and the degree of emphasis in Dutch, but a negative correlation in British English. This is an unexpected difference, which illustrates a case of ambiguity in the meaning of pitch.

## 1. Introduction

Two previous studies on the universality and language-dependence of intonational meanings derived from Ohala's Frequency Code [1, 2, 3] have shown that although it is universally used, speakers of different languages vary in degree to which these meanings are perceived [4, 5]. The Frequency Code is biologically determined and based on the fact that smaller larynxes tend to produce higher notes than larger ones. One of its manifestations is that high pitch sounds vulnerable and submissive, while low pitch sounds protective and dominant. In [5], two affective meanings of the Frequency Code were assessed in Dutch and British English (henceforth BrE), as used on utterances representing various speech acts and realized with %L H*L L% and %L L*H H% intonation contours. (For ToDI transcription conventions, see <www.lands.let.kun.nl/todi>.) Utterances identical in lexical content and phonological pitch contour differed from each other in pitch range. Pitch range was varied in two ways: (1) pitch span, the difference between the highest F0 and the lowest F0; (2) pitch register, the average F0. It was found that both pitch register and pitch span were relevant to the interpretation of the Frequency Code. By and large, the degree of friendliness increases and the degree of confidence decreases when the pitch range is raised. However, the increase in degree of friendliness is stronger in BrE than in Dutch, while the decrease of confidence is stronger in Dutch than in BrE. Largely, at identical pitch ranges, BrE, as represented in the BrE stimuli, is perceived as more friendly and more confident than Dutch, as represented in the Dutch stimuli. It is well known that Dutch speakers employ a smaller mean pitch range (also referred to as the standard pitch range) than do BrE speakers [e.g. 6]. However, the fact that Dutch listeners differentiated more between stimuli on the confidence scale, but less between stimuli on the friendliness scale, than British English listeners suggested that the difference in the standard pitch range between these two languages does not have a uniform effect on the perception of meanings that are signalled by variations in pitch height.

These findings led us to the question how the difference in the standard pitch range would affect the use of the Effort Code [7,8]. The Effort Code is based on the fact that greater articulatory effort tends to create more elaborate and more explicit phonetic realisations. In the case of fundamental frequency variation, greater explicitness leads to more canonical, wider pitch movements, less explicitness to slurred, narrow-range movements. As a result, a wider pitch span is associated with meanings that can be derived from speakers' motivations for the expenditure of articulatory effort. Two types of interpretations of the biological codes were suggested by [7], informational and affective interpretations. The informational interpretations of the Effort Code is 'emphatic', derived from the perception that the speaker regards his or her message as important and thus spends more energy on its production. An affective interpretation is 'surprised', derived from the perception that the speaker shows agitation, thereby spending more effort on speech production.

It is assumed that in the pitch domain, greater articulatory effort is most noticeable in the pitch span of a stretch of speech. A larger pitch span is often achieved by raising the high values. Therefore, peak height is a relevant variable for the manifestation of the Effort Code. As delayed peaks can be used as substitutes for higher peaks and, like higher peaks, can signal interrogativity [4,7], later peaks may also be related to meanings derived from the Effort Code and thus be used to signal 'emphasis' and 'surprise'. In addition, great articulatory effort might lead to a higher pitch register, or at least, higher registers may be perceived as being intended to signal higher values for meanings derived from the Effort Code. Hence, peak height, peak alignment and pitch register may all be relevant to the interpretation of meanings derived from the Effort Code. We formulate our research questions as follows:

(1) Do Dutch and BrE listeners differ in their interpretation of meanings derived from the Effort Code, as signalled by peak height, peak alignment and pitch register?
(2) Are any such differences attributable to the degree in which listeners differentiate between F0 steps for each of these variables?

Because we expected the Effort Code to be widely used for different communicative purposes across languages, we thought that the difference between any two languages would most probably be a matter of degree. That is, we expected to find only positive correlations between perceived degree of meanings and each of the three variables, in both languages, even though Dutch listeners might perceive a larger difference between the lowest and the highest stimuli than BrE listeners. Because speakers of Dutch habitually use a smaller pitch range than speakers of BrE, Dutch listeners might expect

Dutch speakers to be more parsimonious with their pitch excursions than BrE listeners would expect speakers of BrE to be, given equal communicative needs. As a result, Dutch listeners might, for instance, interpret a given pitch range as signalling more emphasis than BrE listeners. The expectation of finding this difference is fed by the different standard pitch ranges of the two languages. We label our predictions as the Relative Scale and illustrate it by means of Figure 1. By the Relative Scale, a given F0 is perceived as ranking higher on a 'surprise' scale and an 'emphasis' scale by listeners with a narrow-range language background than by listeners with a wide-range language background. Inasmuch as later peaks can substitute for higher peaks, peak alignment may be subject to a similar difference.

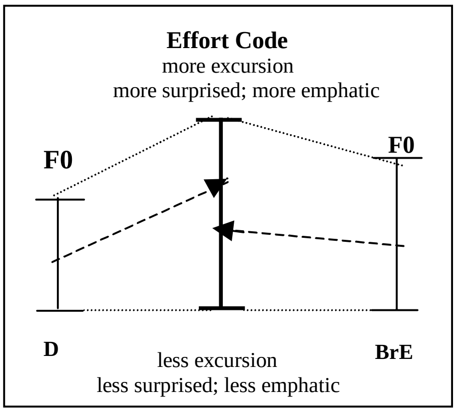


*Figure 1. The relative projection of a given F0 value on the Effort Code scale (middle scale). The scale on the left stands for the standard pitch range of Dutch and the scale on the right stands for the standard pitch range of BrE. The dotted lines indicate how listeners project the standard pitch range on to the Effort Code Scale. The dashed lines indicate the degree to which meanings derived from the Effort Code are perceived for a given F0 value in each language.*

## 2. Perception experiment

In order to find out how Dutch and BrE listeners differ in their interpretation of two meanings derived from the Effort Code, two equivalent perception experiments were carried out in Dutch and BrE. In these experiments, native speakers from the two languages were asked to listen to a number of sentences in their native language and judge these on the semantic scales EMPHATIC vs. NOT EMPHATIC and SURPRISED vs. NOT SURPRISED.

### 2.1 Stimuli

Since the difference in sentence mode may interact with the meaning of intonation contours, we designed three pairs of sentences, where the members of each pair differed in that one was a syntactic question and the other a syntactic statement, but were otherwise identical. Each sentence was composed of three constituents: Subject, Predicate, (Indirect Object) and Object; and a single sentence accent was assigned to the Object.

Natural productions of these six sentences in each language served as the source utterances for the stimuli. These were read by a female Dutch-BrE bilingual speaker. The recording was done in the studio of the Arts Faculty of the University of Nijmegen. Selected readings of the six sentences were digitized at a 32 KHz sampling rate. Speech manipulation was performed by means of the PSOLA technique [8].

In order to study the effect of each of three prosodic variables in a context where the effects of two variables can be measured, but the effect of the third is controlled for, we designed three sets of stimuli, in which two variables were varied and everything else was fixed. These three stimulus sets were the Peak Height-T% (HT) set, the Alignment-Peak Height (AH) set, the Alignment-T% (AT) set. In a fourth set, the overall Pitch Register (PR) was varied. Each source utterance was assigned the contour %L H*L T%, where T% varied between H% and L% in the case of the first three stimulus sets, and was set at H% in the PR set. H* was realized as a 60 ms high plateau, preceded by a 150 ms rise and followed by a 150 ms fall. The high plateau started from the CV boundary of the vowel of the accented syllable. Other pitch points were fixed as indicated in Figure 2.

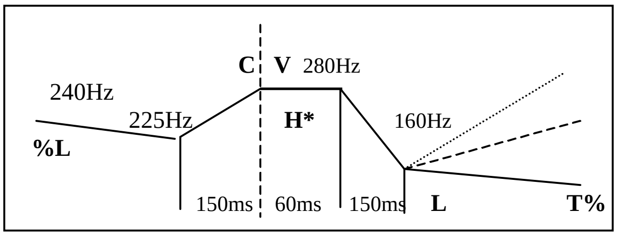


*Figure 2. Stylised contour for %L H*L T%*

In the HT set, peak height was varied from 280 Hz to 400 Hz in steps of 30 Hz, while L remained 160 Hz throughout. While working on the stimuli, we noticed that identical peak heights sounded lower when followed by a rising boundary tone than when followed by a falling boundary. We conjectured that a rising end pitch might mask the effects of peak height. In order to minimize the effects of boundary tones, we included a low T% (130 Hz), a medium T% (280 Hz) and a higher T% (360 Hz), whereby the first of these would be interpreted as L% and the latter two as H%. Each of the five peak height values was combined with each of the three end pitch values. These 15 combinations of peak height and end pitch were distributed over six source utterances, giving us 90 HT stimuli.

In the AH set, each of the five peak height values was combined with each of three peak alignments by shifting the rise-plateau-fall contour through the vowel of the accented syllable in two 50 ms-steps, starting from a position with the onset of the high plateau at the CV boundary. End pitch was fixed at 130 Hz throughout. These 15 combinations of peak alignment and end pitch were distributed over six source utterances, giving us 90 AH stimuli.

Later peaks may be used as substitutes for higher peaks [7], and rising end pitch might therefore mask the effects of a later peak, just as it might a higher peak. In order to minimize the effect of end pitch, each of the three end pitch values was combined with three peak alignments by shifting the rise-plateau-fall contour in the same way as for the AH stimuli. Peak height remained constant at 310 Hz throughout. These nine combinations of peak alignment and end pitch were distributed over six source utterances, giving us 54 AT stimuli.

Finally, in the PR set, each source utterance was assigned the contour %L H*L H%, in which %L, L, and H% were varied in five steps of 20 Hz and H* was varied in 5 steps of 30 Hz. The timings and the lowest F0 values of these points were fixed, as indicated in Figure 2, except that the lowest value of H% is 280Hz. The five pitch register values were distributed over six source utterances, giving us 30 PR stimuli.

Because in 66 cases, the combination of these variables yielded identical stimuli, which were only included once in the total set, we arrived at 90+90+54+30=264, minus 66, or 198 experimental stimuli in each language. Because of the diversity of the stimuli, we thought it was not necessary to include fillers. However, we added a practice session including eight stimuli generated from sentences other than the six source utterances read by the same speaker. These 8 stimuli were also included as the end-of-the-list stimuli, two of them being used twice. The experimental stimuli and the end-of-the-list were mixed manually and then divided into 13 blocks of 16 stimuli in each language. The manual mixture and division were identical for the two languages. To provide listeners with a point of orientation on the scale, we added an anchor stimulus to the beginning of each block of 16 stimuli. There was a 4.5 s pause between stimuli, and a 7 s pause between blocks. Each block was preceded by a 300 Hz sine wave warning tone. In order to minimize order effects, we prepared two stimulus orders by randomizing each block internally and the blocks as a whole twice. Randomizations were identical for the two languages. Each of the two stimulus orders was recorded onto a digital audiotape and then copied to a TDK audiotape. This gave us two 24-minute stimulus tapes in each language.

In addition, we chose to use the magnitude estimation method [9] to obtain the perceptual judgments, as in [4, 5].

### 2.2 Procedure

Twenty-six linguistically naïve native speakers of Dutch (11 man and 15 woman) and 26 of BrE (7 man and 19 woman) between 18 and 30 years old took part in the experiment in equivalent circumstances. During the first session of the experiment, subjects were instructed to try to imagine themselves as the addressees of each stimulus and to indicate the degree to which the speaker expressed surprise, and to record their judgment on the scale 'SURPRISED vs. NOT SURPRISED'. During the second session of the experiment, subjects were instructed to indicate how emphatically each sentence was said by the speaker, and to record their judgment on the scale 'EMPHATIC vs. NOT EMPHATIC'.

## 3. Results

Four sets of data were obtained from the four sets of stimuli, consisting of perceived surprise scores and perceived emphasis scores. Analyses of Variance [10] were performed on each set of data for each of the two dependent variables, i.e. the two semantic scales. These eight Analyses of Variance comprised the between-subject factor Language (2 levels) and the within-subject factor Sentence Mode (2 levels), in addition to the variables that were varied in each data set. As we are mainly interested in the effects of Peak Height, Peak Alignment, and Pitch Register on the perception of the meanings derived from the Effort Code as a function of Language, we will here only consider significant interactions between the between-subject factor Language and the within-subject factors Peak Height, Peak Alignment, and Pitch Register, for each semantic scale. We adopt a significance level of 0.05 and report p-values after correction for sphericity (Huynh-Feldt ε).

### 3.1 The scale EMPHATIC vs. NOT EMPHATIC

With regard to the informational meaning 'emphatic', we found significant interactions between Peak Height and Language in both the HT set ($F_{4,192}$=47.41, $p<0.05$) and the AH set ($F_{4,192}$=12.31, $p<0.05$), as well as between Pitch Register and Language ($F_{4,192}$=15.16, $p<0.05$) in the PR set.

The two-way interaction of Peak Height × Language shows that there is a positive correlation between peak height and the degree of emphasis. This is an evident manifestation of the Effort Code. However, at identical peak heights, Dutch stimuli are perceived as more emphatic than the BrE ones, as illustrated in Figure 3. This difference in degree of the use of the Effort Code is in agreement with the hypothesis of the relative scale as given in Figure 1. We would suggest that this difference is related to the smaller standard pitch range of Dutch. Assuming that the communicative needs of Dutch and BrE speakers are identical where the expression of emphasis is concerned, the smaller space that Dutch speakers allow themselves will need to be exploited more intensively, with the result that a given F0 value is perceived as expressing a higher degree of emphasis by Dutch listeners than by BrE listeners.

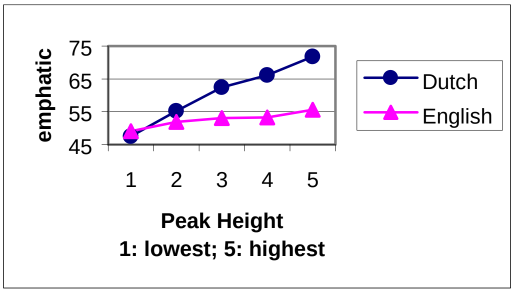


*Figure3. The interaction of Peak Height × Language in the HT set*

An apparent violation of the general hypothesis that the meanings derived from the Effort Code, like those of the Frequency Code and the Production Code, are universal [4,7,11], was found in the two-way interaction of Pitch Register × Language: it reveals that when the pitch register is raised, the degree of emphasis increases for Dutch listeners, but decreases for BrE listeners, as illustrated in Figure 4.

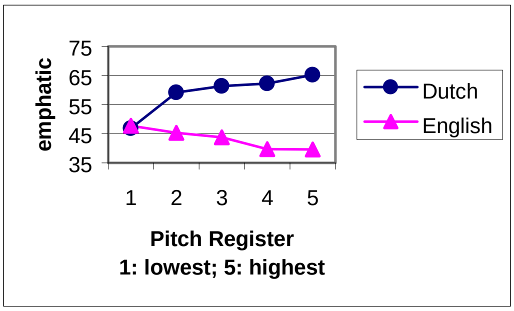


*Figure 4. The interaction of Pitch Register × Language*

This result need not be interpreted to mean that the meanings derived from the biological codes are not universal. Pitch register is not the primary way in which variation along the meaningful scales derived from the Effort Code is expressed. The primary correlate is pitch excursion size, i.e. pitch span. Our suggestion is that Dutch listeners are happy to take high register as a signal for wide span, but that BrE listeners do not show this inclination. This in turn suggests that BrE listeners expect variation in pitch register to be used for the expression of a different meaning, in particular one that Dutch listeners are less inclined to perceive. We suggest that this is precisely the friendliness meaning, which is derived from the Frequency Code. An earlier experiment [5] revealed that Dutch listeners perceive much smaller friendliness differences between one register and the next than BrE listeners.

### 3.2 The scale SURPRISED vs. NOT SURPRISED

With regard to the affective meaning 'surprise', there were significant interactions between Peak Height and Language in both the HT set ($F_{4,192}$=7.22, $p<0.05$) and the AH set ($F_{4,192}$=5.49, $p<0.05$), between Alignment and Language ($F_{2,96}$=3.87, $p<0.05$) in the AH set, and between Pitch Register and Language ($F_{4,192}$=4.37, $p<0.05$ ) in the PR set.

The two-way interaction of Peak Height × Language is similar to that found for the scale EMPHATIC vs. NOT EMPHATIC. In both languages, there is a positive correlation between peak height and the degree of surprise. The meaning surprise can be interpreted as being derived from the Effort Code, through being associated with the expenditure of effort as a result of the state of agitation of the speaker. The interaction is due to the fact that at identical peak heights, the Dutch stimuli are perceived as more surprised than the BrE ones. This difference can again be attributed to the smaller standard pitch range of Dutch, quite analogously to the situation for perceived emphasis.

The two-way interaction of Alignment × Language shows that there is a positive correlation between peak alignment and the degree of surprise in both languages. This is again a manifestation of the Effort Code: a later peak is a substitute for a higher peak [7] and therefore signals a higher degree of surprise, just as would a higher peak. At identical alignments, Dutch stimuli are perceived as more surprised than BrE stimuli, as shown in Figure 5. This difference can be accounted for by referring to the smaller standard pitch range of Dutch.

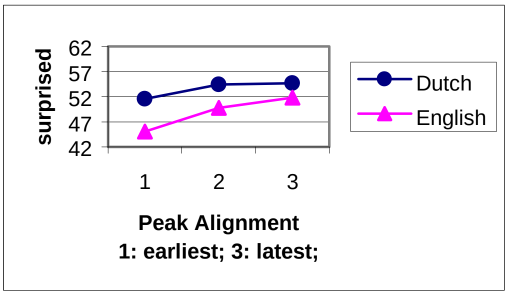


*Figure 5. The Interaction of Alignment × Language.*

Finally, the two-way interaction of Pitch Register × Language shows that in both languages, pitch register is positively correlated with the degree of surprise. This is again a manifestation of the Effort Code. However, again, the meaning differences heard by Dutch listeners are larger than those heard by BrE listeners, if F0 differences are identical.

## 4. Conclusions

In their perception of 'emphasis' and 'surprise', Dutch and BrE listeners behave differently. Two types of language-dependence were found: (1) a difference in the degree to which the meanings are perceived as a function of variation in pitch span and pitch register; (2) a difference in the direction of the correlation between perceived meaning and pitch register. The second was an unexpected type of language-dependence.

Differences in the degree to which phonetic variation of variables that determine pitch span and pitch register lead to differences in the perception of 'surprise' and 'emphasis' in Dutch and BrE can be understood as being due to the smaller standard pitch range of Dutch, if we were to find that Dutch listeners perceive larger meaning differences for the same F0 intervals than BrE listeners. This is in fact what we did find, and we anticipated this result, labelling it the relative scale effect [5], shown in Figure 1. The idea is that Dutch speakers allow themselves less F0 space to express the same range of information (degrees of emphasis, degree of surprise) than BrE speakers, and that listeners adjust their interpretative policies accordingly. Both meanings are derived from the Effort Code, which holds that larger excursions stand for meanings that are naturally associated with the expenditure of effort, such as the significance of the message (emphasis, insistence) or state of agitation of the speaker (surprise).

In particular, the variable peak height correlated with greater perceived emphasis and surprise in both language groups, but led to larger perceived differences in the Dutch group. As for pitch register, the same result was found in the case of perceived surprise: the higher the register, the higher the surprise, but in the Dutch group, the perceived differences between the highest and the lowest stimuli were larger.

The second type of difference between the Dutch and BrE results was that in one case, perceived emphasis as a function of pitch register, there was a positive correlation for the Dutch listeners, but a negative correlation for the BrE listeners. It is difficult to see how this result can be explained on the basis of the difference in standard pitch range. At first site, it also goes against the working hypothesis of our investigations that meanings expressed in the phonetic implementation of intonation contours are a reflection of three universally recognised biological codes, the Frequency Code, the Effort Code and the Production Code [7,11]. However, ambiguity is inherent in a situation where three codes are used that employ a single phonetic parameter, fundamental frequency. In particular, high pitch registers can be interpreted as signalling submissiveness and friendliness by the Frequency Code, through the correlation of small size and high pitch [1,2,3], the opposite of assertiveness and authority. Equally, it can signal insistence and emphasis, through the correlation between greater effort and larger excursions, if the added assumption is made that high pitch can be a used as a cue for wide pitch span. The Dutch listeners appear to use the latter interpretation, the BrE listeners the former. That is, in neither case are we dealing with arbitrary meaning attributions to pitch variation, and Dutch and BrE thus both exhibit universal meanings of intonation at the level of phonetic implementation.

## 5. References


[1] Ohala, J.J., 1983. Cross-language use of pitch: an ethological view. *Phonetica* 40, 1-18.

[2] Ohala, J.J., 1984. An ethological perspective on common cross language utilization of F0 in voice. *Phonetica* 41, 1-16.

[3] Ohala, J.J., 1994. The frequency code underlines the sound symbolic use of voice of pitch. In *Sound symbolism,*, L. Hinton et al. (ed.). Cambridge: Cambridge University Press, 325-347.

[4] Gussenhoven, C.; Chen, A.J., 2000. Universal and language-specific effects in the perception of question intonation. *Proceedings of the 6th ICSLP*. 91-94.

[5] Chen, A.J.; Rietveld, A.; Gussenhoven, C., 2001. Language-specific Effects of Pitch Range on the Perception of Universal Intonational Meaning. *Proceedings of the 9th Europspeech.* 1403-1406.

[6] de Pijper, J., 1983. *Modelling British English Intonation*. Dordrecht: Foris Publications.

[7] Gussenhoven, C., Forthcoming. Why question intonation rise and why they sometimes fall. In *Questions Conference*, M. Hoey; N. Scott (ed.). Liverpool: University of Liverpool Press.

[8] Boersma, P.; Weenink, D. Praat: doing phonetics by computer. http://www.fon.hum.uva.nl/praat/

[9] Zraick R.; Liss, J.M., 2000. A comparison of equal-appearing interval scaling and direct magnitude estimation of nasal voice quality. *Journal of Speech, Language, and Hearing Research* 43, 979-988.

[10] Rietveld, A.; van Hout, R., 1993. *Statistical techniques for the study of language and language behaviour*. Berlin: Mouton de Gruyter.

[11] Gussenhoven, C., forthcoming. *The Phonology of Tone and Intonation*. Cambridge: Cambridge University Press.



**Acknowledgements**: We would like to thank Hedy Kamara for assistance with the recording, Joe Niski, Margit Aufterbeck and colleagues from the Phonetics Laboratory of Cambridge University for their help with the administration of the tests, Henriette Hendriks for providing the first author with a roof during her stay in Cambridge, Joop Kerkhoff for allowing us to freely call upon his programming expertise and Mikel Kors for his useful feedback as our first listener.