

# Language-Specificity in the Perception of Paralinguistic Intonational Meaning

Aoju Chen<sup>1</sup>, Carlos Gussenhoven<sup>2</sup>,  
and Toni Rietveld<sup>3</sup>

<sup>1</sup> Max Planck Institute for Psycholinguistics

<sup>2</sup> Queen Mary, University of London

<sup>3</sup> Radboud University Nijmegen

## Key words

*gradience*

*intonational meaning*

*language-specificity*

*perception*

*pitch range*

## Abstract

This study examines the perception of paralinguistic intonational meanings deriving from Ohala's Frequency Code (Experiment 1) and Gussenhoven's Effort Code (Experiment 2) in British English and Dutch. Native speakers of British English and Dutch listened to a number of stimuli in their native language and judged each stimulus on four semantic scales deriving from these two codes: SELF-CONFIDENT versus NOT SELF-CONFIDENT, FRIENDLY versus NOT FRIENDLY (Frequency Code); SURPRISED versus NOT SURPRISED, and EMPHATIC versus NOT EMPHATIC (Effort Code). The stimuli, which were lexically equivalent across the two languages, differed in pitch contour, pitch register and pitch span in Experiment 1, and in pitch register, peak height, peak alignment and end pitch in Experiment 2. Contrary to the traditional view that the paralinguistic usage of intonation is similar across languages, it was found that British English and Dutch listeners differed considerably in the perception of "confident," "friendly," "emphatic," and "surprised." The present findings support a theory of paralinguistic meaning based on the universality of biological codes, which however acknowledges a language-specific component in the implementation of these codes.

*Acknowledgments:* We would like to thank Vincent van Heuven, Jill House and Patricia Vermillion for useful discussions on pitch range and valuable suggestions on the experimental design, Hedy Kamara for her assistance with the recording, Eric Kellerman for his help with producing the answer sheets, Monique van der Haagen, Joe Niski, Gerard Latjes, Patricia Vermillion, Brechtje Post, Margit Aufterbeck, Rachel Smith for their help with recruiting subjects, Francis Nolan for allowing us to make use of the facilities of the Phonetics Laboratory at Cambridge University, Geoff Potter for his technical assistance, Joop Kerkhoff, Rob Kerkhoff, Ellen Kerkhoff, Sofie Kienhorst, Jean-Pierre Sprengers for their assistance with processing the data, and Frauke Hellwig for making the figures and tables. We would also like to thank Patricia Vermillion, Brechtje Post, and Henriëtte Hendriks for kindly hosting the first author during her stays in London and Cambridge. Finally, we are extremely grateful to Mary Beckman, Bob Ladd, Marc Swerts, Vincent van Heuven, and an anonymous reviewer, whose constructive criticisms have enabled us to improve this paper greatly.

*Address for correspondence:* Aoju Chen, Max Planck Institute for Psycholinguistics, P.O. Box 310, 6500 AH Nijmegen, The Netherlands. Fax: +31 (0) 24 3521213; e-mail: <aoju.chen@mpi.nl>.

## 1 Introduction

It is generally accepted that paralinguistic uses of intonation are remarkably similar across languages. That is, while languages can and do differ widely in how discrete pitch categories are used to contrast lexical or phrasal meanings, all languages seem to use gradient variations in pitch properties to signal different degrees of a certain set of meanings. These meanings range from attitude or emotional state of the speaker (e.g., friendliness, confidence, surprise etc.) to attributes concerning the message (e.g., certainty, continuation, and finality).

In line with Ohala's (1983, 1984) view that the uses of pitch in human communication have an ethological basis, Gussenhoven (2002) proposed that all paralinguistic uses of pitch stem from three biologically determined conditions that cause pitch variation within or across speakers. First, smaller larynxes produce higher-pitched sounds because they contain lighter and smaller vocal cords, with which faster vibration rates can be achieved for a given amount of muscular energy. Second, speech production requires energy and a change in the amount of energy expended will be reflected in the speech signal. Third, the supply of this energy is only available in phases, as determined by the breathing process. These three conditions and their corresponding interpretations are referred to as the Frequency Code (Ohala, 1983, 1984), the Effort Code, and the Production Code (Gussenhoven, 2002). In speech production, communication by means of these codes does not require the actual physiological conditions to be created; "it is enough to create the effects" (Gussenhoven, 2002). In speech perception, it can be reflected in that listeners are capable of interpreting pitch variations in the speech of the others in accordance with these codes. The specific meanings that speakers/listeners attach to these codes include attributes of the message (informational meanings) and the speaker (affective meanings). In the next three paragraphs, the three biological codes will be considered in detail.

The Frequency Code was initially proposed by Ohala to account for the cross-language uses of pitch on the basis of the widely attested functions of avian and mammalian vocalizations in hostile situations (Morton, 1977). Emphasizing the fact that the larynx varies in size across speakers, leading to differences in pitch between men and women and between children and adults, Gussenhoven pointed out, following Ohala, that the correlation between larynx size and vibration rates of the vocal cords is exploited for the expression of power relations, such that a speaker can signal a "small" meaning by means of a higher pitch and a "big" meaning by means of a lower pitch. The informational interpretations include "uncertain" (for a higher pitch) versus "certain" (for a lower pitch). The affective interpretations include "feminine," "submissive," "friendly," "polite," and "vulnerable" for a higher pitch and "masculine," "dominant," "confident," "protective," and "aggressive" for a lower pitch.

Greater articulatory effort tends to lead to greater articulatory precision (de Jong, 1995). In the context of intonation, greater precision means less slurring together of pitch movements and less undershooting of tonal targets, which cause them to be carried out with wider excursions. The Effort Code thus associates wider pitch excursions with meanings that can spring from the speaker's motivations for the expenditure

of articulatory effort. The informational interpretations include “emphatic” and “significant”: the speaker is being forceful because he considers his message important. Affective interpretations include “surprised” and “agitated.” A less common interpretation is “obliging”: “going to some lengths in realizing pitch movements may be indicative of an obliging disposition” (Gussenhoven, 2002).

The relevant biological condition for the Production Code is that the generation of the subglottal air pressure required for the vibration of the vocal cords is associated with the exhalation phase of the breathing process. Towards the end of an exhalation phase, there is a fall-off of the subglottal air pressure. Assuming there is a correlation between utterances and exhalation phases, the Production Code associates high pitch with utterance beginnings and low pitch with utterance endings. Thus, high beginnings signal new topics, low beginnings continuation of topics. A reverse relation holds for utterance endings: “high endings signal continuation, low endings finality and end of turn” (Gussenhoven, 2002).

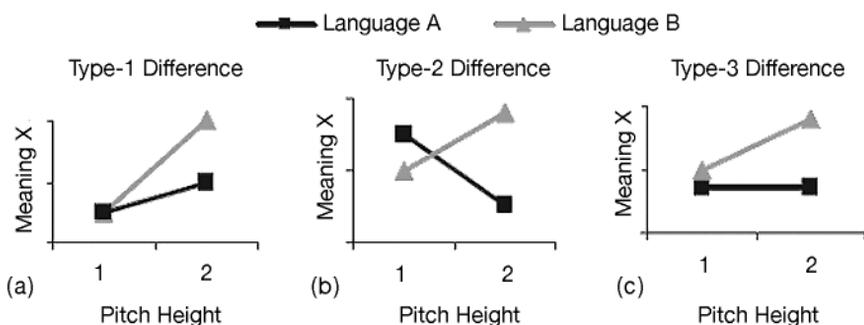
Given that the paralinguistic usage of pitch is biologically determined, it is natural to expect that paralinguistic intonational meaning is completely universal. However, there are indications that this may not be the case. Research on vocal expression of emotion and recognition of emotion (e.g., Scherer, 2000; van Bezooijen, 1984) has shown that although there are universal vocal cues for emotion, including the use of pitch, cultural-specific variations exist. Cross-linguistic investigation of the perception of question (Gussenhoven & Chen, 2000; Hadding-Koch & Studdert-Kennedy, 1964; Makarova, 2000a, 2000b) has demonstrated that although listeners with different language backgrounds appear to rely on a similar set of intonational cues when judging whether an utterance is a question, they differ in their sensitivity to these cues. In this study, with the aim to further establish the nature of paralinguistic intonational meaning, we addressed the question of whether the perception of paralinguistic intonational meaning is the same in different speech communities and as the biological codes predict.

Theoretically, listeners of any two languages, say Language A and Language B, can differ in at least three ways in their perception of meanings deriving from the biological codes. These three types of language-specificity are illustrated in Figure 1. Figure 1a shows a Type 1 difference. Although an increase in pitch is perceived to signal a higher degree of Meaning X by both groups of listeners, the increase in perceived Meaning X is stronger for listeners from Language B than for listeners from Language A. In other words, listeners from Language B use the biological code at issue more intensively than listeners from Language A such that for a given pitch interval, they perceive a larger meaning distinction. This is the sort of difference that might well occur between listener groups whose languages differ in standard pitch range (i.e., the mean pitch range of a language within which its speakers habitually speak), assuming that speech communities with different standard pitch ranges express meaning to the same degree. By contrast, a Type 2 difference, shown in Figure 1b, would be unexpected, since the two groups of listeners perceive reversed form-function relations between pitch variation and Meaning X. In this case, raising the pitch increases the perceived degree of Meaning X in Language B, but decreases it in Language A. Presumably, only one of these form-function relations conforms to the biological code. Finally, a Type 3 difference is shown in Figure 1c. This would

also be unexpected. An increase in pitch leads to an increase in the perceived degree of Meaning X in Language B but no change in Language A. As in a Type 2 difference, the two groups of listeners differ in that one group perceives the meaning as the biological code predicts but the other does not.

**Figure 1**

Three types of hypothetical language-specificity in the perception of paralinguistic intonational meanings deriving from the biological codes. The y-axis indicates scale values for Meaning X, increasing from the lower end to the upper end of the axis. The x-axis indicates pitch variations with Level 2 standing for a higher pitch



Considering that the effects consistent with the biological codes appear to be widely observed across languages (Gussenhoven, 2004), we thought it unlikely that a Type 2 difference or a Type 3 difference could be found between any two languages. The existence of Type 1 differences is suggested by the fact that languages do in fact differ in standard pitch range. A case in point is that British English has a wider standard pitch range than Dutch. De Pijper (1983) found that a pitch range of 12 semitones was appropriate for the realizations of the pitch contours in his British English intonational model, whereas for Dutch a pitch range of six semitones had earlier been adopted ('t Hart & Cohen, 1973). Willems (1982) reported that the range of the falling pitch movement varied between one semitone and seven semitones among Dutch speakers but between four semitones and 12 semitones among British English speakers when reading English declarative sentences. Dutch learners of English performed significantly better in imitating British English pitch patterns after receiving the instruction to increase their pitch range by lowering the low tones and increasing the high tones than without it (de Bot, 1982).

The idea that the standard pitch range can influence the perception of intonational meaning was first put forward by Rietveld, Gussenhoven, Wichmann, and Grabe (1999). In a study on the communicative effects of rising and falling pitch contours in British English and Dutch, Rietveld et al. (1999) assessed a number of attributes of pitch contours  $H^*L L\%$  and  $L^*H H\%$ , as used on utterances representing speech acts including ANSWER, INSTRUCTION, REQUEST and *wh*-QUESTION. It was found that in the speech acts ANSWER and INSTRUCTION,  $L^*H H\%$  was evaluated more negatively than  $H^*L L\%$  by Dutch listeners only, whereas in the speech acts REQUEST and

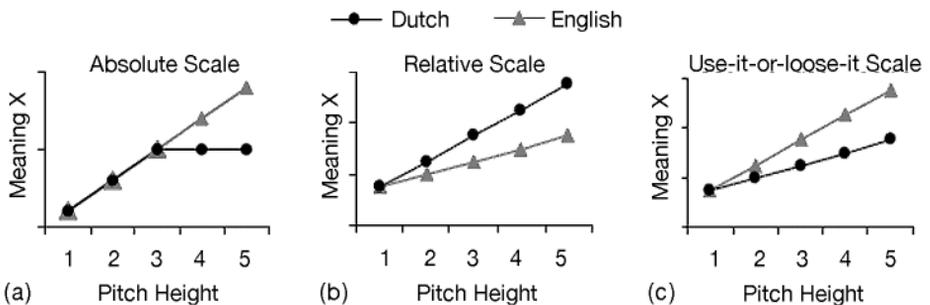
*wh*-QUESTION, L\*H H% was evaluated more positively by British English listeners only. In their perception experiment, the same pitch range was used for English and Dutch stimuli. The authors accounted for these unexpected differences between British English and Dutch as resulting from a difference in standard pitch range. Specifically, they claimed that the negative effect of L\*H H% in speech acts ANSWER and INSTRUCTION was not present in British English stimuli because the pitch excursion was not wide enough for British English listeners to trigger the meaning “challenging” (cf. Gussenhoven, 1984), while the positive effect of L\*H H% in the speech acts REQUEST and *wh*-QUESTION was not experienced by Dutch listeners because L\*H H% stimuli simply sounded too high to be heard as “friendly.”

The Frequency Code and the Effort Code may be more sensitive to standard pitch range than the Production Code, because they involve global as well as local pitch variation (see Section 2.1 for further discussion), whereas the Production Code mainly involves pitch variation at the beginning and end of an utterance.<sup>1</sup> Thus, in order to get a better insight into how speech communities differ in the perception of paralinguistic intonational meaning, we investigated the perception of meanings embodied in the Frequency/Effort Codes in British English and Dutch.

For the reasons discussed above, we concerned ourselves with the question of whether there were Type 1 differences to be found between British English and Dutch listeners. A priori, there are different ways in which the interaction between language and meaning could play out within a Type 1 difference, based on how speakers project their pitch range onto a given semantic scale. We hypothesized three situations, as illustrated in Figure 2.

**Figure 2**

Three implementations of a Type 1 difference in the perception of paralinguistic intonational meanings deriving from the biological codes: the y-axis indicates scale values for Meaning X, increasing from the lower end to the upper end of the axis; the x-axis indicates pitch height variation, increasing from Level 1 to Level 5



<sup>1</sup> However, Swerts, Bouwhuis and Collier (1996) argued that span and register can contribute to the perception of finality as well.

The situation in Figure 2a is referred to as the Absolute Scale. In this situation, British English and Dutch speakers project their pitch range in the same way on the semantic scale, such that a pitch value will correspond with identical semantic values in the two languages. The only difference is due to the fact that British English speakers employ higher pitch values and can thus express greater degrees of meaning. In either language, when pitch values fall out of the standard pitch range, listeners will fail to perceive any further meaning differences, and presumably “out-of-range” pitch values (here Levels 4 and 5 are supposed to be out-of-range in Dutch) would be perceived to be semantically equivalent to the highest pitch value within the standard range. In Figure 2b we give the more plausible Relative Scale, which assumes that both British English and Dutch speakers express the same semantic range, in spite of their different standard pitch ranges. As a result, a given pitch interval will be perceived to signal a larger meaning difference in Dutch than in British English. In Figure 2c, we give the Use-it-or-lose-it Scale. Here, the narrow-range language Dutch fails to use the biological codes at all intensively. The idea here is that since there is not much meaning to be expressed by means of pitch variation, there is no point in having a wide standard pitch range in the first place. By contrast, the wide-range language, British English, has a wide range exactly because its speakers are keen to express the meanings encoded in the biological codes. As a result, a given pitch interval will be perceived to signal a smaller meaning difference in Dutch than in British English.

We hypothesized that any differences in the perception of paralinguistic intonational meanings by British English and Dutch listeners would be classifiable as Type 1 differences with the Relative Scale. To keep the study to a controllable length, the perception of four meanings was examined, “confident” and “friendly,” affective interpretations of the Frequency Code, “surprised,” an affective interpretation of the Effort code, and “emphatic,” an informational interpretation of the Effort Code. These meanings were chosen because they were considered representative of the two biological codes and readily interpretable for speakers with different language backgrounds. More specifically, our hypotheses were:

1. Hypothesis 1 (the Frequency Code): Dutch listeners perceive a larger difference in the degree of confidence and friendliness for a given pitch interval than British English listeners;
2. Hypothesis 2 (the Effort Code): Dutch listeners perceive a larger difference in the degree of surprise and emphasis for a given pitch interval than British English listeners.

Accordingly, two cross-linguistic perception experiments, Experiment 1 and Experiment 2, were carried out in British English and Dutch, in which Hypothesis 1 and Hypothesis 2 were tested by means of four semantic scales: SELF-CONFIDENT versus NOT SELF-CONFIDENT, FRIENDLY versus NOT FRIENDLY (Experiment 1); SURPRISED versus NOT SURPRISED, and EMPHATIC versus NOT EMPHATIC (Experiment 2).<sup>2</sup>

## 2 General methodological issues

Several aspects of our method are common to both experiments. To begin with, since contextual factors such as speaker-listener relationship, lexical content of the preceding utterance, and the intonation of the preceding utterance may interact with the meaning of pitch contours (Cruttenden, 1997), we used individual, lexically unsurprising sentences containing a single accented word as the source utterances for the stimuli. Secondly, the source utterances were recorded in Southern Standard British English and Standard Dutch by a female early British English and Dutch bilingual on DAT tape (48 kHz in 16 bits) in the sound-attenuated studio of the Faculty of Arts at the Radboud University Nijmegen. For each source utterance, the reading with the best sound quality (i.e., clear articulation, modest intensity, average speaking rate, and no creaky voice) was digitized at a 32 kHz sampling rate and subjected to speech manipulation, which was performed by means of Praat (Boersma & Weenink, 1996) using the PSOLA technique. It may be added that using unintelligible speech (i.e., low-pass filtered stimuli) appears to lead to fuzzier results with smaller effect sizes. In a pilot study, we compared the perception of question intonation of an unknown language by Dutch listeners who heard the low-pass filtered stimuli and that by a different group of Dutch listeners, who received the stimuli in the normal speech mode. Although the same tendencies were observed in the perceptual judgments of both groups of listeners, the effect sizes of the significant interactions and main effects were noticeably smaller in the data obtained from the “low-pass filter” group. Thirdly, we chose to vary pitch of the source utterances in terms of hertz instead of semitones or ERB; the scales are virtually the same in the range between 150 Hz and 320 Hz (Rietveld & van Heuven, 1997), which approximated the pitch range adopted for our experiments.

Six other common aspects that deserve more attention will be described in detail in what follows, including pitch range variation, test tapes, subjects, experimental procedure, the Visual Analog Scale—the device used to obtain perceptual judgments, and data processing.

### 2.1

#### ***Pitch range variation***

Pitch range can be varied in two ways, span and register (Cruttenden, 1997, p. 47, p. 123; Ladd, 1996, p. 260). Pitch span reflects the distance between the highs and the lows of the pitch contour. Pitch register reflects the mean pitch value of the pitch contour (Rietveld & Vermillion, 2003).<sup>3</sup> Of the two types of pitch variation, pitch register variation may be the primary variable employed in the use of the Frequency

<sup>2</sup> The Dutch translations of these scales are as follows: , ZELFVERZEKERD versus NIET ZELFVERZEKERD, VRIENDELIJK versus NIET VRIENDELIJK, VERBAASD versus NIET VERBAASD, NADRUKKELIJK versus NIET NADRUKKELIJK.

<sup>3</sup> See Patterson (2000) for a review of measures of span and register that have been proposed in literature.

Code, as it involves increases or decreases of both the high and low pitches. Pitch span variation will be employed to a lesser extent, because increases in pitch span leave low pitches unaffected, or may lower them even further. For the purposes of the Effort Code, pitch span will be the main variable. However, pitch register may also be involved, because it co-varies with pitch span, that is, the wider the span, the higher the register (Ladd, 1996, p. 260), and a higher pitch register may signal the same meaning as pitch span. Since both pitch span and pitch register may be involved in the expression of the four meanings, we decided to include both as variables. The details on how span and register variations in each experiment were realized are presented in the sections on the experimental design.

## 2.2

### **Test tapes**

Test tapes were prepared containing experimental stimuli, practice trials (Experiment 2) and fillers (Experiment 1). A 4.5 s pause was inserted after each stimulus to allow subjects to give their judgments. There was a 7 s pause between blocks of 16 stimuli and between blocks of eight practice trials. Each block was preceded by a 200 ms 300 Hz sine wave to signal the beginning of the block. An anchor was inserted before the first item of each block. The anchor was a neutral-sounding realization of the utterance *Do you sell ORGANIC fruit as well/Verkoopt u ook BIOLOGISCH fruit* by the same speaker who read the source utterances for the stimuli with the pitch contour %L H\*L H%.<sup>4</sup> %L was fixed as 240 Hz and H% as 320 Hz; the pitch accent H\*L was realized as a 30 ms high plateau (280 Hz) preceded by a 120 ms rise (starting from 225 Hz) and followed by a 120 ms fall (falling to 195 Hz), which was followed by a low-plateau (195 Hz) ending at a point that was 100 ms before the pitch target of H%. The anchor was included to provide listeners with a reference point for their judgments on the semantic scales. It is known that the distance between a given stimulus and the anchor can have an effect on the perceptual judgment. To minimize this effect, the stimuli were randomized such that those representing the same experimental condition could appear at various places in a stimulus block. The randomizations were identical for the two languages and resulted in four stimulus orders in Experiment 1 and two stimulus orders in Experiment 2. Each of the stimulus orders was recorded onto DAT tape (48 kHz in 16 bits) and then copied to a TDK audio tape. This led to four 43 mins test tapes in Experiment 1 and two 24 mins test tapes in Experiment 2 in each language.

## 2.3

### **Subjects**

Two criteria were taken into account when recruiting subjects. As we were not concerned with age-related differences in the perception of intonational meaning, we decided to recruit subjects from a single age range, that is, the range 18–35 years. Since we built our hypotheses on Standard Southern British English and Standard Dutch,

---

<sup>4</sup> Pitch contours are described following the ToDI notation (Gussenhoven, Rietveld & Terken, 1999; Gussenhoven, 2005).

British English subjects were recruited from native speakers of Standard Southern British English and Dutch subjects were recruited from native speakers of Standard Dutch or a variety of Dutch which has not been reported to differ from Standard Dutch intonationally. Our British English subjects were undergraduates from Cambridge University and the University of London. Dutch subjects were students from the Polytechnic College of 's-Hertogenbosch and the Radboud University Nijmegen, and while they spoke standard Dutch, were predominantly from the provinces Noord Brabant and Gelderland. There have been no documented differences in intonation or standard pitch range between the varieties of Dutch spoken in these two provinces and Standard Dutch. Subjects were paid a small fee for their participation.

## 2.4

### **Procedure**

In both languages, the experiment was carried out in two sessions with a five-minute break in between, one session for each semantic scale. Subjects were divided over stimulus orders approximately equally. Because it was not always possible to conduct the experiments in a phonetics/language lab equipped with headphones, we presented stimuli to approximately half of the subjects over headphones in a sound-treated phonetics/language lab and half of the subjects through a Philips AQ6455 cassette recorder/player at an adequate volume in a quiet room. Subjects were instructed by means of written instructions in their native language to try to imagine themselves as the addressees of the stimuli and indicate for each stimulus their impression of the speaker or the stimulus on the semantic scale concerned. No explanation of the meanings of “friendly,” “confident,” “surprised,” and “emphatic” was given in the instructions.<sup>5</sup> Subjects recorded their judgments for each stimulus on a Visual Analog Scale, which is discussed in detail in the following section. At the end of the experiment, subjects were asked to fill out a questionnaire on their language and social background.

## 2.5

### **Visual Analog Scale**

The Visual Analog Scale (VAS) is widely used in clinical research to assess an attribute that is believed to range across a continuum of values and cannot be directly measured, such as pain and mood (Kreiman, Gerratt, Kempster, Erman & Berken, 1993; Wewers & Lowe, 1990). Conventionally, a VAS is a 100 mm line that is either vertically or horizontally oriented, with the two ends anchored to two ends of a given scale, for instance, “no pain” versus “very severe pain.” The patients mark on the line the point that they feel represents their perception of their current state. The VSA score is determined by measuring the millimetres from the left end of the line to the point that the patient marks.

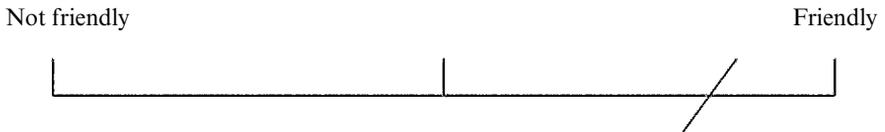
---

<sup>5</sup> A small number of subjects (<10) required an explanation of “emphatic” and/or what “emphatic-sounding” should sound like. In this case, the experimenter asked them to try to recall how they would speak when they had something important to tell the others.

In the present study, horizontal VASs were used to obtain perceptual judgments. The left end of the VAS was the negative end of each semantic scale, that is, “not friendly,” “not self-confident,” “not emphatic,” and “not surprised,” and the right end of the line was the positive end of each semantic scale, that is, “friendly,” “self-confident,” “emphatic,” and “surprised.” The anchor was assigned the score corresponding to the mid-point of the VAS, indicated by a short vertical line. Subjects were instructed to mark on the line the point that represents the perceived degree of the meaning in question for each stimulus by means of slash. The more to the left of the scale they placed the slash, the lower the degree of the meaning they perceived; the more to the right of the scale they placed the slash, the higher the degree of the meaning they perceived. An example of the horizontal VASs is given in Figure 3, in which the listener would have judged the speaker as fairly friendly.

### Figure 3

An example of the Visual Analog Scales used in examining the perception of “friendly”



## 2.6

### **Data processing**

Data containing the VAS scores for a given semantic attribute were obtained by measuring the distance of the slash from the left end in terms of millimetres. The score 1 was given if the slash was placed exactly at the left end of the VAS. A placement of the slash exactly at the right end of the VAS was given the score 99. Placing the slash at the midpoint of the VAS was given the score 50. Missing values were noted separately.

Before the data were statistically analyzed, a data-trimming procedure was conducted to exclude data obtained from subjects with any one of the following characteristics: (1) early bilinguals, defined as bilinguals who acquired two languages before the age of 12; (2) having lived in countries other than the Netherlands (regarding Dutch subjects) or the U.K. (regarding British English subjects) before the age of 12; (3) speaking varieties of English/Dutch that differ substantially from Standard Southern British English and Standard Dutch; (4) falling outside the chosen age range; (5) not having performed well in the task. Poor performance was interpreted as leaving 5% or more of the stimuli unjudged with respect to one or both of the semantic scales and/or judging all the utterances as neutral-sounding irrespective of the semantic scales. The exact number of subjects whose data were excluded will be specified per characteristic in the sections on statistical analyses and results.

### 3 Experiment 1: Language-specificity in the use of the frequency code

#### 3.1

##### *Experimental design*

On the assumption that speaker-listener relationship may interact with the interpretation of how friendly or confident the speaker sounds, speech act was included as a within-subject factor. Three speech acts were taken into account, that is, Information, Request, and Instruction, implemented as four *wh*-questions, four yes-no questions, and four non-question sentences (3 statements and 1 imperative) respectively. These 12 sentences (see Appendix 1) served as the source utterances for the stimuli. Examples of the source utterances are given in English and Dutch in (1). The accented syllables are in capitals.

1. Information: What's the L<sub>Level</sub> of this course? / Wat is het ni<sub>VEAU</sub> van deze cursus?
- Request: Could you please carry that T<sub>ABLE</sub> down the stairs? / Zou je de TA<sub>fel</sub> even naar beneden kunnen dragen?
- Instruction: You should fill out the CL<sub>AIM</sub> form. / Je moet het decla<sub>RAT</sub>ie formulier invullen.

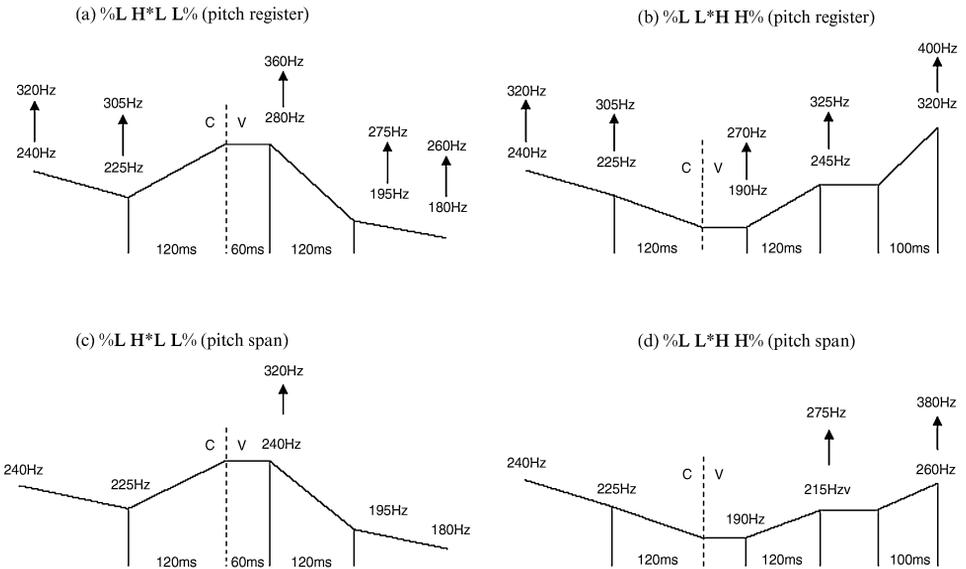
Pitch contour was also included as a within-subject factor because it interacts with speech act and hence can influence the perception of intonational meaning. Each source utterance was assigned two contours, %L H\*L L% (abbreviated H\*L L%) and %L L\*H H% (abbreviated L\*H H%). The chosen pitch contours are naturally oriented towards different ends of the Frequency Code, with the rising contour L\*H H% associating with the “small” meanings and the falling contour H\*L L% with the “big” meanings. The schematic representations of H\*L L% and L\*H H% are given in Figure 4. In H\*L L%, the H\*L pitch accent was realized as a 60ms high plateau preceded by a 120ms rise and followed by a 120ms fall. In L\*H H%, the L\*H pitch accent was realized as a low plateau covering the full length of the accented vowel, preceded by a 120ms fall and followed by a 120ms rise, which was followed by a high plateau (or by a subsequent further rise). The 60ms high plateau of H\*L and the low plateau of full vowel length of L\*H started at the CV boundary of the accented syllable, but 10ms after the CV boundary of the accented syllable if it had a voiced coda or was followed by at least another syllable, and 10ms before the CV boundary if it had a consonant cluster as onset or was preceded by at least another syllable (Rietveld & Gussenhoven, 1995). The high plateau of L\*H was of varying duration and ended at a point that was 100ms before the pitch target of H%; it was not realized when the sonorant material was sparse in the segments following the accented word (i.e., when the sonorant material could be no longer than 30ms).

Pitch range of the source utterances was varied in terms of pitch register and pitch span. Pitch register was varied in five levels by raising both the H tones (H\*, H, H%) and the L tones (%L, L\*, L, L%) in four equal steps of 20Hz. This gave us 120 stimuli (3 speech acts × 4 sentences × 2 pitch contours × 5 pitch registers) in each

language. We will refer to this set as the Register stimuli. The values used for pitch register variation are given at the two ends of the upper arrows in Figures 4a and 4b. Pitch span was varied in five levels too. Here only the H tones in each contour were varied, as indicated by the upper arrows in Figures 4c and 4d. In H\*L L%, H\* was varied in four equal steps of 20 Hz. In L\*H H%, H was varied in four steps of 15 Hz; H% in four equal steps of 30 Hz.<sup>6</sup> This gave us 120 (3 speech acts × 4 sentences × 2 pitch contours × 5 pitch spans) stimuli varying in pitch span in each language. We will refer to this set as the Span stimuli. Together, the Register and the Span stimuli amounted to 240 experimental stimuli in each language.

### Figure 4

Schematic representations of %L H\*L L% (abbreviated H\*L L%) and %L L\*H H% (abbreviated L\*H H%) with variations in register and span



In order to minimize the risk that subjects would detect a pattern of pitch variation in specific pitch contours, we added 128 fillers, giving a total of 368 stimuli for each language. The source utterances for fillers (see Appendix 2) consisted of three sentences that were used to generate experimental stimuli and six new sentences read by the same speaker and recorded in equivalent circumstances. We included two new pitch contours, %L H\*L H% and %H !H\*L L%, in addition to the two pitch contours used for the experimental stimuli. Values used for pitch range variation were similar to those used for the experimental stimuli but not identical. The experimental stimuli and the fillers were mixed manually and divided into 23 blocks of 16 stimuli.

<sup>6</sup> H and H% were varied in unequal steps such that natural sounding renditions of L\*H H% with five perceptually well distinguishable spans could be generated.

## 3.2

### **Procedure**

Fifty-three linguistically naïve native speakers of Dutch (30 women and 23 men) and 29 of British English (20 women and 9 men) took part in the experiment in equivalent circumstances. Subjects were instructed by means of written instructions in their native language to judge each stimulus on how confident the speaker sounded during one session of the experiment, and how friendly the speaker sounded during the other session of the experiment, and recorded their judgments on the VASs. Approximately half of the subjects started with the “confident” session and the other half of the subjects started with the “friendly” session.

## 3.3

### **Statistical analyses and results**

Two sets of data containing “confidence” scores and “friendliness” scores were obtained from the two sets of stimuli, that is, the Span stimuli and the Register stimuli. Data of 11 Dutch subjects were excluded from statistical analyses, among whom one fell out of the chosen age range, three lived in a foreign country before the age of 12, and seven performed poorly. Data of seven British English subjects were excluded from statistical analyses, among whom two turned out to be from Northern Ireland, and five performed poorly. Thus, data of 42 Dutch subjects (23 women and 19 men) and 22 British English subjects (16 women and 6 men) were subjected to statistical analyses. Repeated measures analysis of variance (ANOVA) was used to analyze the data.

Although VASs measure the magnitude of a given attribute, strictly speaking VAS data are not of the interval level. There is, thus, the concern that the so-called interval statistics like *t*-tests and *F*-ratios in analysis of variance may not be used to analyze the present data. However, many authors (in particular, in the 1960’s and 1970’s) argued that this concern may not be necessary. Harris (1975, p. 226) for instance, stated that “the validity of statistical conclusions depends only on whether the numbers to which they are applied meet the distributional assumptions to derive them, and not the scaling procedures used to obtain the numbers.” It should be added that *t*-tests and *F*-ratios are known to be insensitive to non-normality of the distributions. Harris also pointed out that most scales that are assumed to be of the ordinal interval—at least by a strict interpretation of Stevens’s laws (Stevens, 1946)—contain interval information. This led both Harris (1975) and Labovitz (1970) to quite categorical statements on this matter. The latter (1970, p. 515) is quoted here:

Empirical evidence supports the treatment of ordinal variables as if they conform to interval variables ... Although some small error may accompany the treatment of ordinal variables as interval, this is offset by the use of more powerful, more sensitive, better developed, and more clearly interpretable statistics with known sampling error.

On the basis of these statements—and many more similar statements can be quoted, see for instance Anderson (1961)—we think that analysis of variance can be applied to data that contain information on the magnitude of the differences between objects, as obtained in magnitude scaling and equal appearing interval scaling, even though they are not strictly interval data. This view is also voiced in

more recent work on methodology. For example, Munro and Page (1993) said “there is now sufficient evidence to show that use of parametric tests with ordinal data rarely distorts the results.” Dexter and Chestnut (1995) carried out an interesting simulation experiment, in order to find out whether the analysis of VAS data with parametric or non-parametric tests yielded different results. They found that the *t*-test and ANOVA are good choices to compare VAS measurements among groups.

Four analyses of variance (repeated measures on the within-subject factors) were performed, one on each set of data for each of the two dependent variables, that is, the “confidence” score and the “friendliness” score, at the significance level of .05. Each of these four analyses included one between-subject factor, Language (2 levels), and three within-subject factors, Speech Act (3 levels), Pitch Contour (2 levels), and Pitch Register (5 levels) for the Register data or Pitch Span (5 levels) for the Span data. Missing values were replaced with means of the stimuli from the same condition. Table 1 gives an overview of the results of the analyses.

As we are mainly interested in interactions between the pitch-range related factors and the between-subject factor Language, we will consider significant interactions involving at least Language and one of the pitch-range related factors in detail (see shaded cells in Table 1). If there are related higher-order interactions, two-way interactions will not be discussed. When necessary, significant main effects of the pitch-range related factors are considered. Where relevant, the Huynh-Feldt corrected *p*-values are reported. The measure of effect size used here is partial  $\eta^2$ .

*Perception of “confident”* For the meaning ‘confident’, significant main effects of Pitch Register,  $F(4, 248) = 276.46$ ,  $p < .001$ , partial  $\eta^2 = 0.817$ , and Pitch Span,  $F(4, 248) = 9.66$ ,  $p < .001$ , partial  $\eta^2 = 0.135$ , and the significant interaction of Pitch Register  $\times$  Speech Act  $\times$  Language,  $F(8, 496) = 3.70$ ,  $p < .001$ , partial  $\eta^2 = 0.987$ , will be discussed here.

The main effect of Pitch Register is such that the perceived degree of confidence decreased when the pitch register increased. The main effect of Pitch Span is similar to that of Pitch Register but weaker than we expected: the wider the pitch span, the lower the perceived degree of confidence, as shown in Figure 5. Inspection of the significant two-way interaction of Pitch Span  $\times$  Pitch Contour,  $F(4, 248) = 3.68$ ,  $p < .05$ , partial  $\eta^2 = 0.878$ , reveals that the main effect of pitch span was readily observable in L\*H H% but hardly present in H\*L L%.

Turning to the three-way Pitch Register  $\times$  Speech Act  $\times$  Language interaction, as our concern is language-specificity in the interpretation of pitch register in different conditions, we show the interaction of Pitch Register by Language for each speech act separately in Figure 6. As can be seen, even though both groups of listeners associated a higher pitch register with a lower degree of confidence, Dutch listeners made a noticeably larger meaning distinction for the given interval of pitch registers than English listeners, in particular in speech acts Instruction and Information. Furthermore, in Request, Dutch listeners assigned lower “confidence” scores than English listeners across the five pitch register conditions.

*Perception of “friendly.”* For the meaning ‘friendly’, significant main effects of Pitch Register,  $F(4, 248) = 16.38$ ,  $p < .001$ , partial  $\eta^2 = 0.209$ , and Pitch Span,  $F(4, 248) = 71.79$ ,

**Table 1**

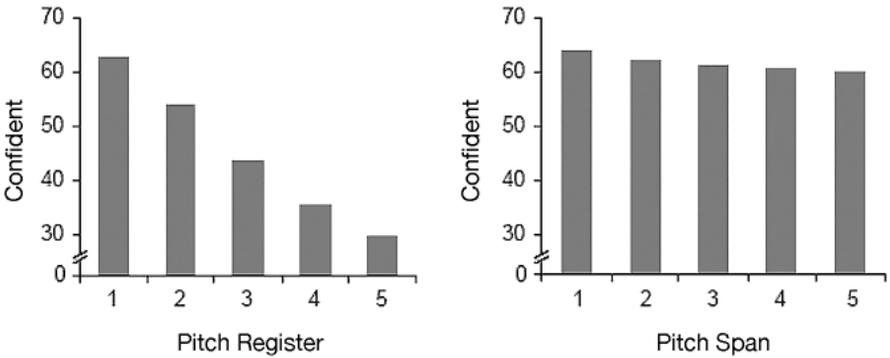
Results of the analyses of variance on the “confidence” score and the “friendliness” score. Significant effects and interactions (.05 level) are marked by ‘\*’; non-significant ones by ‘-’. Empty cells indicate that the effects and interactions are not relevant to the dataset. Shaded cells indicate that the effects and interactions are discussed in detail

	<i>Confident</i>		<i>Friendly</i>	
	<i>Span set</i>	<i>Register set</i>	<i>Span set</i>	<i>Register set</i>
Language (L)	–	–	–	*
Pitch Contour (PC)	*	–	*	*
Pitch Span (PS)	*		*	
Pitch Register (PR)		*		*
Speech Act (SA)	*	*	*	*
PC × SA	*	*	*	–
PS × SA	–		*	
PR × SA		–		*
PC × PS	*		*	
PC × PR		*		*
PC × PS × SA	–		–	
PC × PR × SA		*		–
L × PC	–	–	–	–
L × PS	–		–	
L × PR		–		*
L × SA	–	–	*	*
L × PC × SA	–	–	–	–
L × PS × SA	–		–	
L × PR × SA		*		*
L × PC × PS	–		–	
L × PC × PR		–		–
L × PC × PS × SA	–		–	
L × PC × PR × SA		–		–

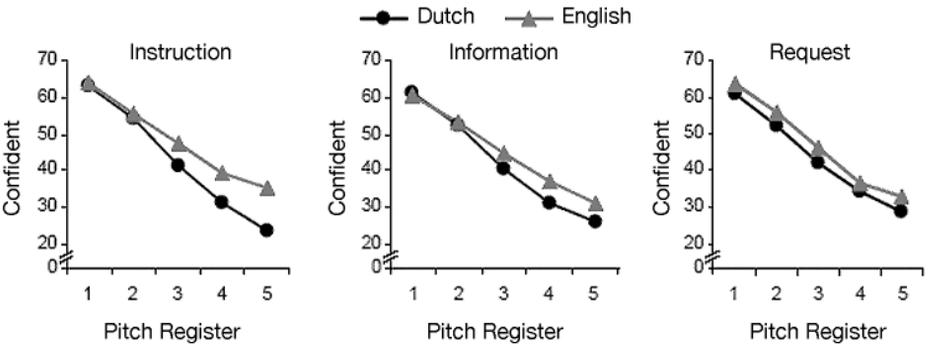
$p < .001$ , partial  $\eta^2 = 0.537$ , and the significant interaction of Language × Pitch Register × Speech Act,  $F(8, 496) = 4.34$ ,  $p < .001$ , partial  $\eta^2 = 0.996$ , will be considered in detail. Figure 7 shows the main effect of Pitch Register and Pitch Span. As can be seen, a larger pitch span was perceived as signaling a higher degree of friendliness. The main effect of Pitch Register is similar to the main effect of Pitch Span: by and large a higher pitch register was assigned a higher “friendliness” score; there is a slight decrease in the perceived degree of friendliness at Level 5.

**Figure 5**

Mean “confidence” scores for all listeners ( $n=64$ ) in five pitch span conditions (panel a) and five pitch register conditions. Pitch Span/Register increases from Level 1 to Level 5

**Figure 6**

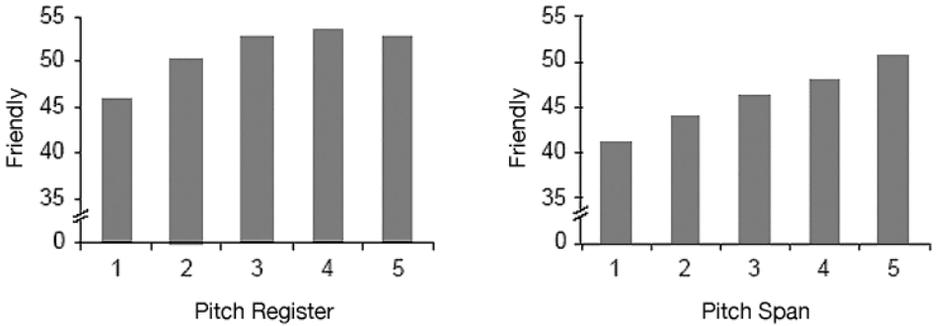
Mean “confidence” scores for Dutch ( $n=42$ ) and English ( $n=22$ ) listeners in five pitch register conditions in the speech acts Instruction, Information and Request. Pitch Register increases from Level 1 to Level 5



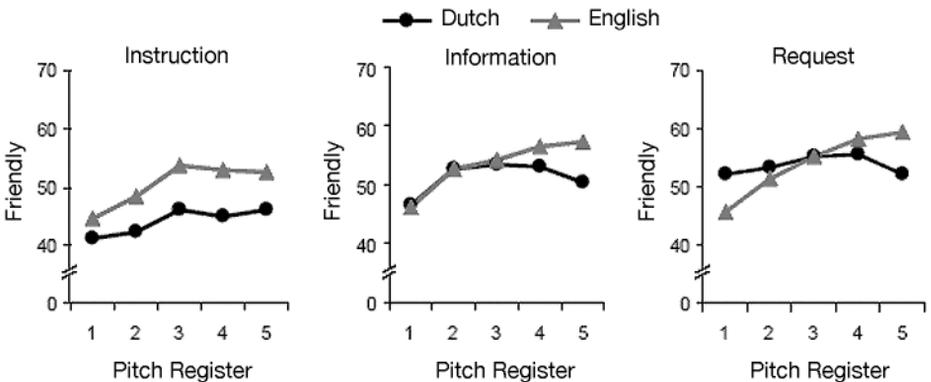
The nature of the three-way interaction of Pitch Register  $\times$  Speech Act  $\times$  Language is shown in Figure 8, which shows the interaction of Pitch Register by Language for each speech act. As can be seen, although by and large an increase in peak height led to an increase in the perceived friendliness in the ratings of both groups of listeners in each speech act, the increase in the perceived friendliness was steeper in English listeners' ratings than in Dutch listeners' ratings, in particular in Information and Request. Note that in Dutch listeners' ratings, there was a slight decrease in the perceived friendliness from Level 4 to Level 5, which explains the decrease at Level 5 observed in Figure 5. Furthermore, in Instruction, the scores were higher in English ratings than in Dutch ratings.

**Figure 7**

Mean “friendliness” scores for all listeners ( $n=64$ ) in five pitch span conditions. Pitch Span increases from Level 1 to level 5

**Figure 8**

Mean “friendliness” scores for Dutch ( $n=42$ ) and English ( $n=22$ ) listeners in five pitch register conditions in the speech acts Instruction, Information and Request. Pitch Register increases from Level 1 to Level 5



### 3.4

#### Discussion

The main effects of Pitch Span and Pitch Register confirm the predictions of the Frequency Code in British English and Dutch: the wider/higher the pitch span/pitch register, the higher the perceived degree of friendliness and the lower the perceived degree of confidence. The effect of pitch span in the perception of “confidence” is, however, weaker than we expected. This may be because a wider span can signal a higher degree of significance according to the Effort Code and the speaker who presents certain information as significant can be seen as an individual with self-confidence. As suggested to us by Mary Beckman, if some listeners interpret a wider span to be less confident sounding (the Frequency Code) and some listeners

interpret a wider span to be more confident sounding (the Effort Code), as found in van Bezooijen (1988), there would be a dilution of the effect.

Unexpectedly, only the interaction of Pitch Register  $\times$  Speech Act  $\times$  Language was found to be significant in respect of each meaning. This indicates that British English and Dutch listeners were similar in the other cases, for instance, in the perception of “confident” and “friendly” as signaled by pitch span variation. With respect to the perception of “confident,” this three-way interaction indicates a Type 1 difference between Dutch and English listeners, in particular, in speech acts Instruction and Information, and can be explained by the Relative Scale. With respect to the language-specificity in the perception of “friendly,” it can be interpreted to indicate a Type 1 difference but cannot be explained by the Relative Scale. According to the Relative Scale, Dutch listeners should make a sharper meaning distinction between the lowest and the highest pitch registers as a result of their smaller standard pitch range. We found exactly the opposite, a result that agrees with the Use-it-or-lose-it scale, which may seem to suggest that Dutch listeners did not use the Frequency Code as intensively as British English listeners, and consequently became less sensitive to it. However, it is not clear why the Use-it-or-lost-it scale should *only* be in effect in the perception of friendliness as signaled by pitch register variation. We postpone further discussion to the next section after the findings obtained from Experiment 2.

## **4 Experiment 2: Language-specificity in the use of the effort code**

### **4.1**

#### ***Experimental design***

Three pairs of sentences were drawn up whose members only differed in the application of subject-verb inversion, creating three minimally different statements and questions. Our motivation for including a variable “sentence type” was that surprise and emphasis may be signaled differently in questions and statements. The six sentences (see Appendix 3) served as the source utterances for the stimuli. Each sentence was composed of Subject, Verb, (optionally Indirect Object) and Object. A single sentence accent was assigned. An example of a sentence pair is given in English and Dutch in (2). The accented syllables are in capitals.

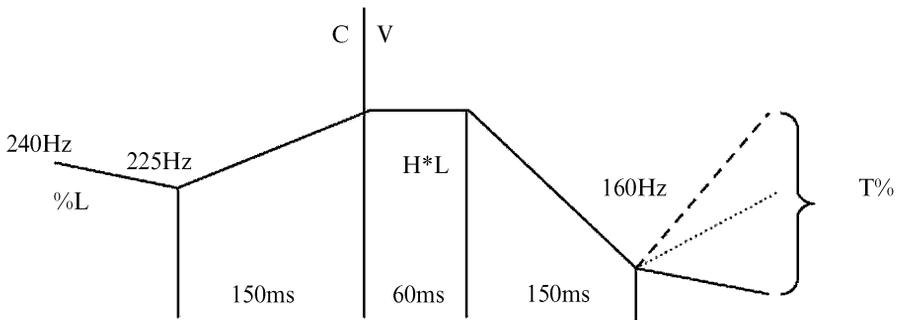
2. Did you ask her for a JOB interview / Heb je haar naar een BAAN gevraagd?  
You asked her for a JOB interview / Je hebt haar naar een BAAN gevraagd.

In addition to Pitch Register and Peak Height (as an alternative to Pitch Span), Peak Alignment and End Pitch were included as variables. Since peak delay can be used as a substitute for, or enhancement of peak raising (Gussenhoven, 1999, 2004), peak alignment might be related to the perception of meanings deriving from the Effort Code.<sup>7</sup> In addition, in the process of speech manipulation, we noticed that identical peak heights sounded lower when followed by a high boundary tone than when followed by a low boundary tone. We earlier found that high end pitches have a masking effect on the prominence of the peak (Chen, 2005, p. 59). End pitch was therefore incorporated as the fourth variable.

To keep the experiment within manageable proportions, we decided to study the effect of each of three prosodic variables: peak height (H), peak alignment (A) and end pitch (T) in combination with one other variable, while the effect of the third was controlled for. Three sets of stimuli were thus designed, in which two variables were varied and everything else was fixed. These three stimulus sets were the Peak Height-End Pitch (HT) set, the Peak Alignment-Peak Height (AH) set, and the Peak Alignment-End Pitch (AT) set. In a fourth set, the overall Pitch Register (PR) was varied. Each source utterance was assigned the contour %L H\*L T%. The boundary tone T% was varied between L% and H% in the case of the HT and AT stimulus sets. It was realized as L% in the AH stimulus set and as H% in the PR stimulus set. The H\*L pitch accent was realized as a high plateau of 60 ms, preceded by a 150 ms rise and followed by a 150 ms fall. As in Experiment 1, the high plateau started at the CV boundary of the accented vowel, but 10 ms after the CV boundary if the accented syllable had a voiced coda or was followed by one or more syllables, and 10 ms before the CV boundary if it was preceded by one or more syllables (Rietveld & Gussenhoven, 1995). Other pitch points were fixed, as indicated in Figure 9a.

### Figure 9a

Schematic representation of the contour %L H\*L T%



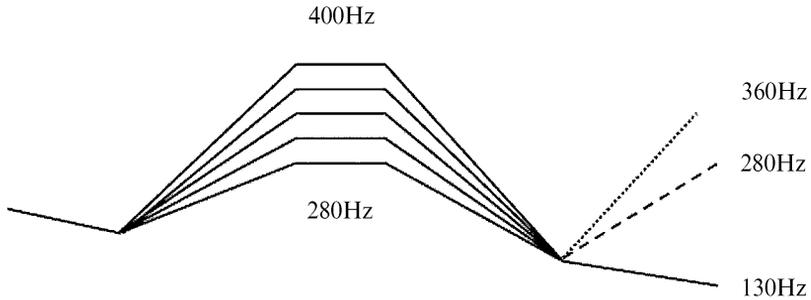
In the HT set, peak height was varied from 280 Hz to 400 Hz in steps of 30 Hz. The boundary tone was varied in three steps, a low T% (130 Hz), a medium T% (280 Hz) and a higher T% (360 Hz), whereby the first of these would be interpreted as L% and the latter two as H%, that is, H% and higher H%, as illustrated in Figure 9b. Note that the pitch interval between L% and H% was made bigger than that between H% and higher H%. This was done to create a clear perceptual difference between L%

<sup>7</sup> The claimed relationship between peak delay and peak raising is based on the observation that a higher pitch peak takes longer to reach than a lower peak given the same rate of pitch change, and therefore tends to be aligned later in the segmental string (Gussenhoven, 2004, p. 90). However, there is evidence suggesting that there is no necessary relationship between these two properties. For example, Braun (2004) found no correlation between peak height and alignment of peak relative to the end of the stressed syllable in the realization of prenuclear accents in German.

and H%. Each of the five peak height values was then combined with each of the three end pitch values, resulting in 15 combinations of peak height and end pitch. These combinations were distributed over the six source utterances, making up 90 HT stimuli.

### Figure 9b

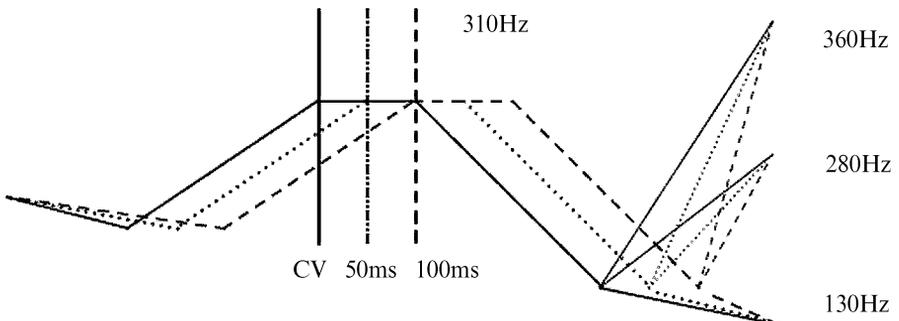
Schematic representation of the contour %L H\*L T% in the Peak Height-End Pitch (HT) stimulus set, with the peak height varied in four equal steps of 30 Hz and end pitch varied in three unequal steps



In the AT set, each of the three end pitch values used in the HT set was combined with three peak alignments by shifting the rise-plateau-fall contour through the vowel of the accented syllable in two 50 ms steps starting at the CV boundary, as illustrated in Figure 9c. Peak height remained constant at 310 Hz throughout. These nine combinations of peak alignment and end pitch were distributed over the six source utterances, forming 54 AT stimuli.

### Figure 9c

Schematic representation of the contour %L H\*L T% in the Peak Alignment-End Pitch (AT) stimulus set, with time alignment of the pitch peak in two 50 ms steps starting from the CV boundary of the accented syllable, and end pitch varied in three unequal steps

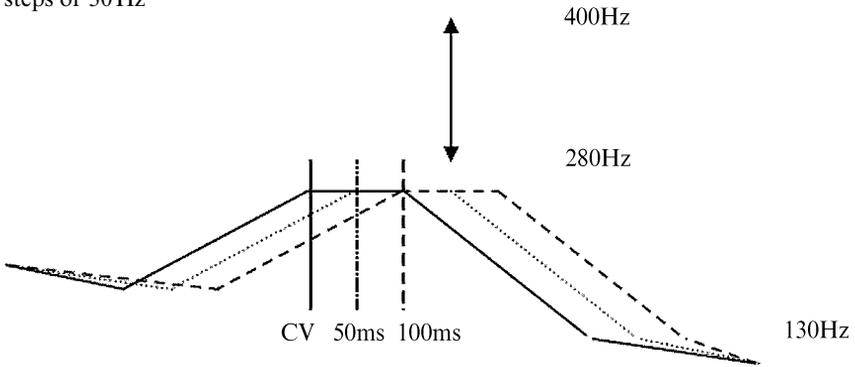


In the AH set, each of the five peak height values used in the HT set was combined with each of three peak alignments by shifting the rise-plateau-fall contour in the

same way as for the AT stimuli, as shown in Figure 9d. End pitch was fixed at 130 Hz throughout. These 15 combinations of peak alignment and peak height were distributed over the six source utterances, making up 90 AH stimuli.

### Figure 9d

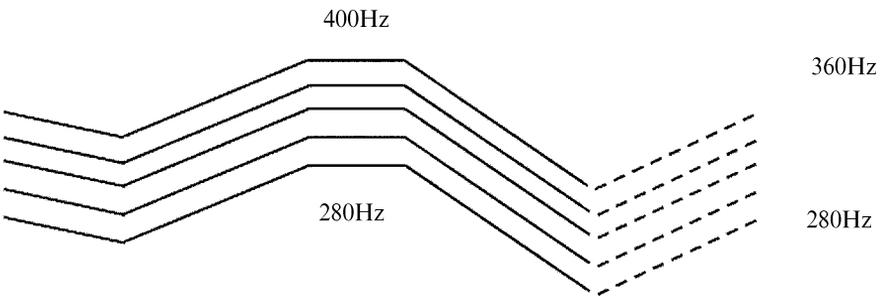
Schematic representation of the contour %L H\* L L% in the Peak Alignment-Peak Height (AH) stimulus set, with time alignment of the pitch peak varied in two 50 ms steps starting from the CV boundary of accented syllable, and the pitch peak of H\* L varied in four equal steps of 30 Hz



Finally, in the Pitch Register set, %L, L, and H% were varied in five steps of 20 Hz and H\* was varied in five steps of 30 Hz, as can be seen in Figure 9e.<sup>8</sup> The time alignments and the lowest F0 values of these points were fixed, as indicated in Figure 9a, except that the lowest value of H% was set at 280 Hz. The five pitch register values were distributed over the six source utterances, making up 30 PR stimuli.

### Figure 9e

Schematic representation of the contour %L H\* L H% in the Pitch Register (PR) stimulus set, with peak height varied in four 30 Hz steps and other pitch points in four 20 Hz steps



<sup>8</sup> It was believed at the time of preparing the stimuli that pitch peak increased in a greater amount than other pitch points when pitch register increased. It may indeed be better to have varied every pitch point in equal steps to achieve variations in pitch register, as suggested by Mary Beckman.

In 66 cases, the combination of these variables yielded identical stimuli. These were only included once in the total set, and the total number of the experimental stimuli amounted to 198 (90 + 90 + 54 + 30 – 66). In addition, eight stimuli were generated from sentences other than the six source utterances read by the same speaker, which served as the practice trials (see Appendix 4). These eight stimuli were also included as the end-of-the-list stimuli, two of them being used twice. The 198 experimental stimuli were mixed manually and divided into 12 blocks of 16 plus one block of six, which formed the last stimulus block together with 10 end-of-the-list stimuli. For each language, there were thus 13 blocks of 16 stimuli, plus a block of eight practice trials.

## **4.2**

### ***Procedure***

Twenty-six linguistically naïve native speakers of Dutch (11 men and 15 women) and 26 of British English (7 men and 19 women) took part in the experiments in equivalent circumstances. Subjects were instructed by means of written instructions in their native language to try to imagine themselves as the addressees of the stimuli, and indicate for each stimulus how emphatically it was said in one session and how surprised the speaker sounded in the other session. Approximately half of the subjects started with the “emphatic” session and the other half of the subjects started with the “surprised” session.

## **4.3**

### ***Statistical analyses and results***

Four sets of data containing “emphasis” scores and “surprise” scores were obtained from the four sets of stimuli. The data of one Dutch subject and one English subject were not included in statistical analyses because of poor performance. Thus, data of 25 Dutch subjects (11 men and 14 women) and 25 British English subjects (7 men and 18 women) were selected for statistical analyses. Separate ANOVAs (repeated measures on the within-subject factors) were performed on each data set for each of the two dependent variables, that is, the “emphasis” score and the “surprise” score, at a significance level of .05. Each analysis included the between-subject factor Language (2 levels) and the within-subject factor Sentence Type (2 levels), in addition to the pitch-related variables in each data set. Missing values were replaced with means of the stimuli from the same conditions. Table 2 gives an overview of the results of the analyses.

As we are mainly interested in how the pitch-range related factors (Peak Height, Peak Alignment, End Pitch and Pitch Register) affect the perception of “emphatic” and “surprised” in British English and Dutch, we will focus primarily on significant interactions involving at least the between-subject factor Language and one of the pitch-range related factors (see shaded cells in Table 2). Generally, a two-way interaction is not considered when there is a related higher-order interaction. When necessary, significant main effects of the pitch-related factors are mentioned. Where relevant, the Huynh-Feldt corrected *p*-values are reported. The measure of effect size used here is partial  $\eta^2$ . Findings concerning Peak Height and Pitch Register are reported and discussed in what follows in this section. A separate section is devoted to findings concerning Peak Alignment and End Pitch.

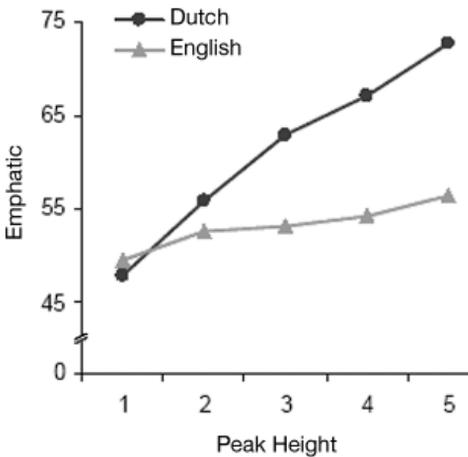
**Table 2**

Results of the analyses of variance on the “emphasis” and “surprise” scores. HT—the Peak Height-End Pitch stimulus set, AT—the Peak Alignment-End Pitch stimulus set, AH—the Peak Alignment-Peak Height stimulus set, and PR—the Pitch Register set. Significant effects and interactions (.05 level) are marked by ‘\*’; non-significant ones by ‘-’. Empty cells indicate that the effects and interactions are not relevant to the dataset. Shaded cells indicate that the effects and interactions are discussed in detail

	<i>Emphatic</i>				<i>Surprised</i>			
	<i>HT</i>	<i>AT</i>	<i>AH</i>	<i>PR</i>	<i>HT</i>	<i>AT</i>	<i>AH</i>	<i>PR</i>
Language (L)	*	-	*	*	-	-	*	-
Peak Height (H)	*		*		*		*	
Peak Alignment (A)		*	-			-	*	
End Pitch (T)	*	-			*	*		
Pitch Register (PR)				*				*
Sentence Type (ST)	*	*	*	*	-	-	*	-
H × ST	-		*		*		-	
A × ST		-	*			*	-	
T × ST	*	-			*	*		
PR × ST				-				*
H × A			-				*	
H × T	-				*			
A × T		-				*		
H × A × ST			-				*	
H × T × ST	-				-			
A × T × ST		-				*		
L × H	*		*		*		*	
L × A		-	-			-	*	
L × T	0.055	-			*	-		
L × PR				*				*
L × ST	-	-	*	-	-	-	*	-
L × H × ST	-		-		-		-	
L × A × ST		-	-			-	-	
L × T × ST	-	-			*	-		
L × PR × ST				*				-
L × H × A			-				-	
L × H × T	-				-			
L × A × T		-				-		
L × H × A × ST			-				*	
L × H × T × ST	-				-			
L × A × T × ST		*				-		

*Perception of “emphatic.”* With respect to the informational meaning ‘emphatic’, the two-way interaction of Peak Height  $\times$  Language in both the HT stimulus set,  $F(4, 192) = 18.76$ ,  $p < .001$ , partial  $\eta^2 = 0.281$ , and the AH stimulus set,  $F(4, 192) = 14.56$ ,  $p < .001$ , partial  $\eta^2 = 0.233$ , and the three-way interaction Pitch Register  $\times$  Sentence Type  $\times$  Language,  $F(4, 192) = 3.771$ ,  $p < .001$ , partial  $\eta^2 = 0.073$ , in the PR stimulus set will be considered in what follows.<sup>9</sup>

The effect of the two-way interaction of Peak Height  $\times$  Language in the HT stimulus and that in the AH stimulus set are nearly identical. Figure 10 shows this two-way interaction as found in the HT stimulus set. As can be seen, for both British English and Dutch listeners, higher peaks were perceived as signaling a higher degree of emphasis. However, the increase in the perceived emphasis across the peak height conditions was clearly steeper in Dutch ratings than in British English ratings.



**Figure 10**

Mean “emphatic” scores for Dutch ( $n = 25$ ) and English ( $n = 25$ ) listeners in five peak height conditions in the Peak Height-End Pitch (HT) stimulus set. Peak Height increases from Level 1 to Level 5

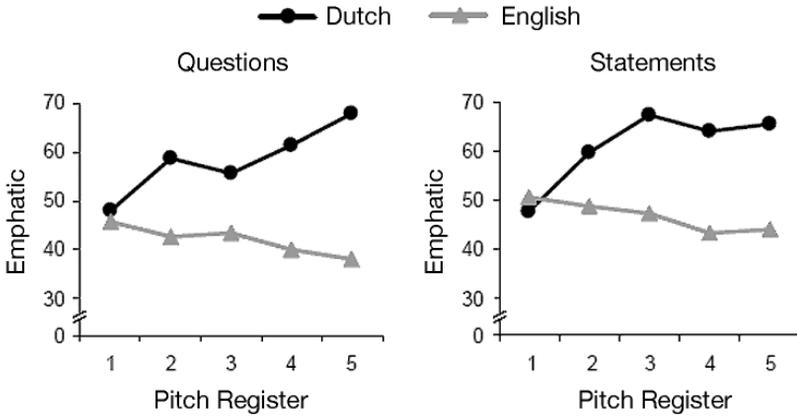
The nature of the interaction of Pitch Register  $\times$  Language is strikingly different from that of the interaction of Peak Height  $\times$  Language. As is evident in Figure 11, when the pitch register was raised, the perceived emphasis largely increased for Dutch listeners but decreased for British English listeners in both questions and statements. Moreover, the increase in the perceived emphasis in Dutch listeners’ ratings is steeper than the decrease in English listeners’ ratings in statements.

*Perception of “surprised.”* With respect to the informational meaning ‘surprised’, the interaction of Peak Height  $\times$  Language in both the HT stimulus set,  $F(4, 192) = 7.066$ ,  $p < .005$ , partial  $\eta^2 = 0.128$ , and the AH stimulus set,  $F(4, 192) = 5.211$ ,  $p < .05$ , partial  $\eta^2 = 0.098$ , and the interaction of Pitch Register  $\times$  Language in the PR stimulus set,  $F(4, 192) = 4.187$ ,  $p < .05$ , partial  $\eta^2 = 0.08$ , will be reported in the following paragraphs.

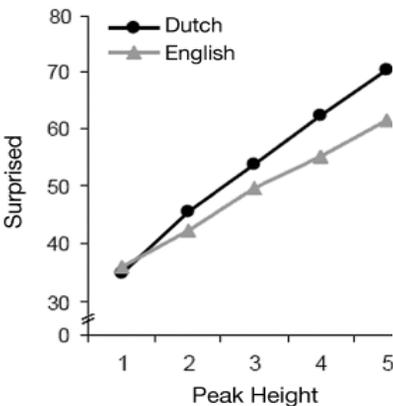
<sup>9</sup> The four-way Peak Alignment  $\times$  End Pitch  $\times$  Sentence Type  $\times$  Language interaction is reported in Table 2, but is not discussed because of its unclear effect on the perception of “emphatic.”

**Figure 11**

Mean “Emphasis” scores for Dutch ( $n=25$ ) and English ( $n=25$ ) listeners in five pitch register conditions in questions and statements in the Pitch Register (PR) stimulus set. Pitch Register increases from Level 1 to Level 5

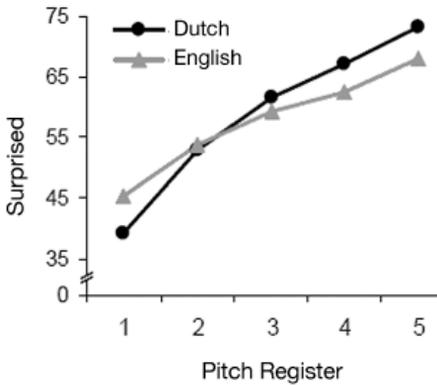


The nature of the two-way interaction of Peak Height  $\times$  Language in the HT stimulus set is similar to that found in the AH stimulus set, which is shown in Figure 12. For both British English and Dutch listeners, a higher peak signaled a higher degree of surprise. However, the increase in the perceived surprise was somewhat steeper in Dutch ratings than in British English ratings. To further determine whether the difference between the two groups of listeners was significant at Levels 2, 3, 5, and 5, simple contrasts were obtained for Peak Height, for which Level 1 was used as the control category. The contrast between Dutch listeners' and British English listeners' ratings was significant at Level 2,  $F(1, 48) = 5.425, p < .05$ , partial  $\eta^2 = 0.102$ ; Level 3,  $F(1, 48) = 4.554, p < .05$ , partial  $\eta^2 = 0.135$ ; Level 4,  $F(1, 48) = 7.084, p < .05$ ; and Level 5,  $F(1, 48) = 7.479, p < .05$ , partial  $\eta^2 = 0.128$ , compared to at Level 1. This result indicates that the difference in the increase in the perceived surprise from Level 1 to a higher level of peak height was significant between Dutch and British English ratings.

**Figure 12**

Mean “surprise” scores for Dutch ( $n=25$ ) and English ( $n=25$ ) listeners in five peak height conditions in the Peak Alignment-Peak Height (AH) stimulus set. Peak Height increases from Level 1 to Level 5

The two-way interaction of Pitch Register  $\times$  Language is shown in Figure 13. As can be seen, in both languages, pitch register was positively correlated with the degree of surprise. However, again, the increase in the perceived surprise was somewhat steeper in Dutch ratings than in British English ratings. To further determine the significance of this difference, simple contrasts were requested for Pitch Register, for which Level 1 was used as the control category. The contrast between Dutch listeners' and British English listeners' scores was significant at Level 2,  $F(1, 48) = 4.165$ ,  $p < .05$ , partial  $\eta^2 = 0.08$ ; Level 3,  $F(1, 48) = 4.902$ ,  $p < .05$ , partial  $\eta^2 = 0.093$ ; Level 4,  $F(1, 48) = 5.533$ ,  $p < .05$ , partial  $\eta^2 = 0.103$ ; and Level 5,  $F(1, 48) = 6.194$ ,  $p < .05$ , partial  $\eta^2 = 0.114$ , when using Level 1 as a reference. This result indicates that the difference in the increase in the perceived surprise from Level 1 to a higher level of pitch register was significant between Dutch and British English ratings.



**Figure 13**

Mean "surprise" scores for Dutch ( $n = 25$ ) and English ( $n = 25$ ) listeners in five pitch register conditions in the Pitch Register (PR) stimulus set. Pitch Register increases from Level 1 to Level 5

#### 4.4

##### **Discussion**

Taken together, the differences between British English and Dutch listeners in the perception of emphasis as signaled by peak height variation and the perception of surprise as signaled by both peak height and pitch register variation are clearly Type 1 differences. They can be explained by the Relative Scale. Hypothesis 2 (i.e., Dutch listeners perceive a larger difference in the degree of surprise and emphasis for a given pitch interval than British English listeners) is thus largely borne out.

Interestingly enough, in the perception of emphasis as signaled by pitch register variation, British English listeners and Dutch listeners opted for reversed interpretations of the form-function relation between pitch register and the meaning 'emphatic'. For Dutch listeners, the higher the pitch register, the higher the perceived degree of emphasis; whereas for British English listeners, the higher the pitch register, the lower the perceived degree of emphasis.<sup>10</sup> Clearly, this is a Type 2 difference. It is difficult

<sup>10</sup> Because this difference between the two groups of listeners was not present in the perception of emphasis as signaled by pitch span variation, this finding supports our assumption that pitch register affects the use of the Effort Code, partially independently of pitch span. A similar observation was made with respect to the meanings "afraid" and "confident" in Patterson and Ladd (1999).

to see how this result can be explained on the basis of the difference in standard pitch range. At first sight, it also goes against the working hypothesis of our investigation that the form-function relations between pitch variation and meaning are reflections of biological codes, since these codes imply that the form-meaning relations derived from them are universally recognized.

However, ambiguity is inherent in a situation where three codes employ a single phonetic parameter, that is, F0. In particular, high pitch register can be interpreted to signal submissiveness and friendliness by the Frequency Code. Equally, it can signal emphasis and significance through the correlation between greater effort and larger pitch movement. Dutch listeners appeared to use the latter interpretation, but British English listeners the former. As the meaning ‘emphatic’ and the meaning ‘friendly’ undermine each other, so to speak, British English listeners associated an increase in pitch register with an increase in the perceived friendliness but a decrease in the perceived emphasis. If this explanation were correct, we would expect British English listeners to make a sharper meaning distinction for a given interval of pitch register in the perception of friendliness than Dutch listeners, despite the wider standard pitch range of British English. As reported in Figure 8, this is indeed what we found. We therefore attribute the Type 2 difference in the perception of emphasis as signaled by pitch register variation and the Type 1 difference in the perception of friendliness as signaled by pitch register variation to a difference in choice as to which biological code is being addressed with increases in register, rejecting the Use-it-or-lose-it account.<sup>11</sup>

## 4.5

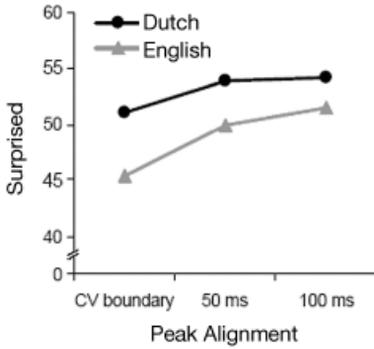
### **Language-specific effects of Peak Alignment and End Pitch**

Peak Alignment and End Pitch were incorporated in the experimental design as secondary variables of pitch range variation because of their relevance to the perception of peak height. In this connection, it is interesting to report that language-specific effects were also found for these secondary variables in the perception of the meaning ‘surprised’ and arguably the meaning ‘emphatic’. The following significant interactions are relevant to illustrating this point and will be discussed in detail. The last interaction is concerned with the meaning ‘emphatic’; the other interactions are concerned with the meaning ‘surprised’.

- Alignment × Language in the AH set,  $F(2, 96) = 3.48, p < .05$ , partial  $\eta^2 = 0.068$ ;
- Alignment × Peak Height × Sentence Type × Language in the AH set,  $F(8, 384) = 2.7, p < .05$ , partial  $\eta^2 = 0.053$ ;
- End Pitch × Sentence Type × Language in the HT set,  $F(2, 96) = 5.21, p < .05$ , partial  $\eta^2 = 0.098$ ;
- End Pitch × Language in the HT set,  $F(2, 96) = 3.61, p = .055$ , partial  $\eta^2 = 0.07$ , in the perception of “emphatic.”

<sup>11</sup> This also accounts for the decrease in the perceived friendliness at Levels 4 and 5 of Pitch Register in Dutch listeners’ ratings (Fig. 8). That is, these high registers might have sounded so emphatic to Dutch listeners that they stopped perceiving an increase in friendliness.

The effect of the two-way interaction of Alignment  $\times$  Language is shown in Figure 14. As can be seen, there is a steady increase in the perceived surprise across the three levels of Peak Alignment in English listeners' ratings but only from Level 1 to Level 2 in Dutch listeners' ratings. Simple contrasts were calculated for Peak Alignment, for which Level 1 was used as the control category. The contrast between Dutch listeners' and English listeners' ratings was significant at Level 3,  $F(1, 48) = 5.613, p < .05$ , partial  $\eta^2 = 0.105$ , compared to at Level 1. This result indicates that the increase in the perceived surprise from Level 1 to Level 3 of Peak Alignment found for Dutch listeners did not hold for English listeners. Arguably, this difference may be interpreted as a weak version of a Type 3 difference.



**Figure 14**

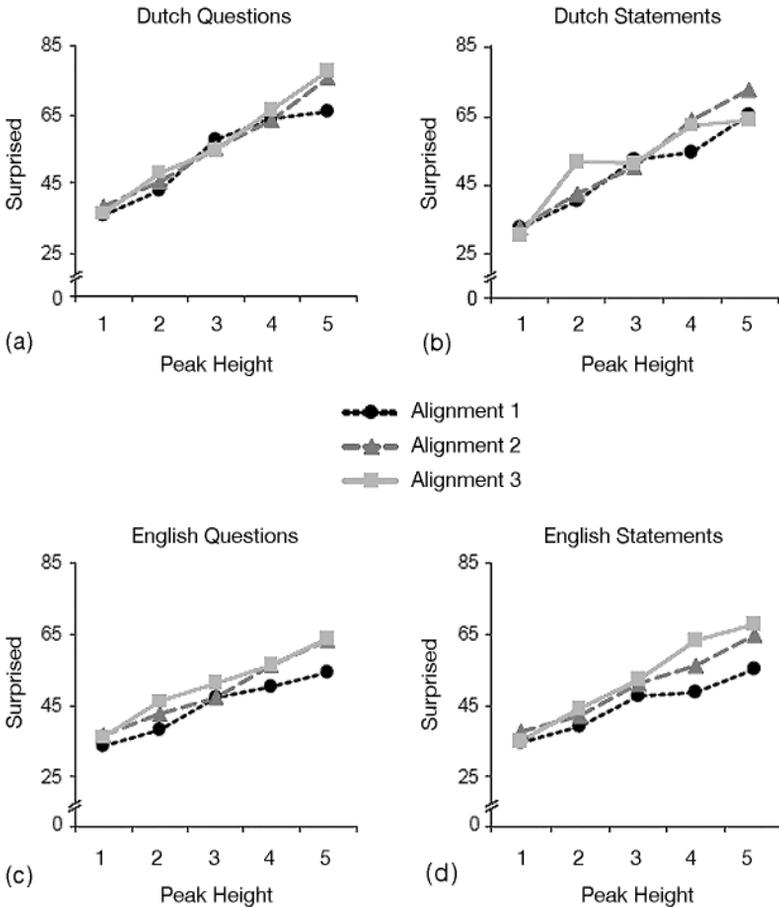
Mean “surprise” scores for Dutch ( $n = 25$ ) and English ( $n = 25$ ) listeners in three peak alignment conditions in the Peak Alignment-Peak Height (AH) stimulus set. At Level 1, the onset of the high plateau was aligned with the CV boundary of the accented syllable; at Levels 2 and 3 the high plateau started at 50 ms and 100 ms to the right the CV boundary respectively

It is not clear how we should account for this difference, but we might speculate that Dutch listeners/speakers are not inclined to exploit the secondary variable peak alignment whereas English listeners/speakers are. This seems to be supported by the four-way Peak Alignment  $\times$  Peak Height  $\times$  Sentence Type  $\times$  Language interaction found in the same stimulus set. As a later peak can function as an enhancement of a higher peak, equally high peaks would sound more surprising when aligned later than when aligned earlier. This pattern is largely borne out in English listeners' ratings, as is evident in Figures 15c and 15d. By contrast, identical peak heights were perceived to signal a similar degree of surprise across the three peak alignment conditions by Dutch listeners (Figs. 15a and 15b). English listeners' willingness to exploit the phonetic space of peak alignment in the signaling of “surprise” is probably rooted in the fact that peak alignment signals meaning along the continuum of “routineness” in British English: “the degree of non-routineness and the degree of delay increase proportionately” (Gussenhoven, 1984, p. 219). There is also evidence for exploitation of peak alignment in American English, coming from the rise-fall-rise contour with its two variants, L + H\* L-H% and L\*+H L-H% (Hirschberg & Ward, 1992; Ward & Hirschberg, 1985). These two contours are said to differ mainly in the alignment of pitch peak or the timing of the first rise. The peak in L\*+H L-H% is aligned later than that in L + H\* L-H% and indicates “incredulity” (about a scale implicitly invoked by the hearer as a response to what the speaker previously said) or “uncertainty” (about a scale invoked by the speaker himself). By contrast, a contour with a delayed peak is a less commonly reported contour in Dutch. ToDI (Gussenhoven, 2005; Gussenhoven, Rietveld, & Terken, 1999) has one, that is, the delayed fall (L\*HL); the IPO grammar

(Collier and 't Hart, 1981; 't Hart, 1998; 't Hart, Collier & Cohen, 1990) has 3C, comparable to L\*HL, which is often used in Dutch TV commercials.

**Figure 15**

Mean “surprise” scores of Dutch listeners ( $n=25$ ) and English listeners ( $n=25$ ) in the five peak height conditions across three peak alignment conditions in questions and statements in the Peak Alignment-Peak Height (AH) stimulus set. Peak Height increase from Level 1 to Level 5. Peak delay increases from Level 1 to Level 3

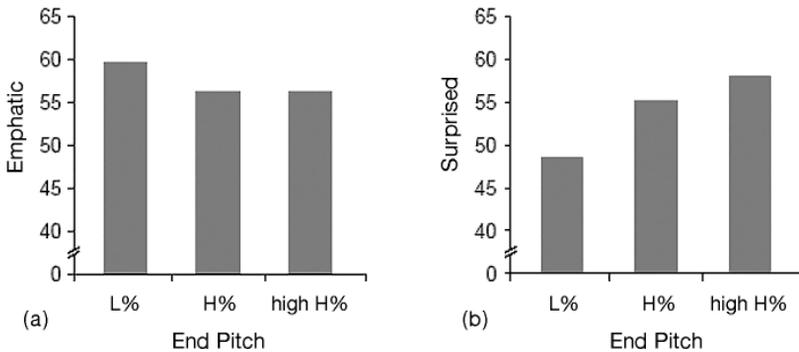


Before considering the language-specific effects of End Pitch, it is useful to inspect the main effect, which was found to be significant for the meaning ‘emphatic’ in the HT stimulus set,  $F(2, 96) = 4.75, p < .05$ , partial  $\eta^2 = 0.09$ , and for the meaning ‘surprised’ in both the HT stimulus set,  $F(2, 96) = 42.38, p < .001$ , partial  $\eta^2 = 0.469$ , and the AT stimulus set,  $F(2, 96) = 24.55, p < .001$ , partial  $\eta^2 = 0.338$ . Figure 16a shows the main effect of End Pitch in respect of the meaning ‘emphatic’: there was an observable decrease in the perceived emphasis from Level 1 to Level 2. The main effect

of End Pitch in respect of the meaning ‘surprised’ is radically different. There is a demonstrable increase in the perceived surprise from Level 1 to Level 2 and a relatively smaller increase from Level 2 to Level 3, as can be seen in Figure 16b, which shows the effect of End Pitch as found in the HT stimulus set.<sup>12</sup> Because Level 1 of End Pitch is a phonological L% while Levels 2 and 3 are phonetic variants of H%, the main effect of End Pitch can be interpreted to mean that L% and H% have different effects on the perception of emphasis than on the perception of surprise. That is, L% triggers the perception of a higher degree of emphasis, whereas H% triggers the perception of a higher degree of surprise, with a higher H% sounding more surprised. In these cases, the distribution of the form over the function is not predicted by the Effort Code, which merely associates wide excursion with higher degree of emphasis and surprise, but has nothing to say about the direction of pitch movements.

**Figure 16**

Mean “emphasis” scores (panel a) and mean “surprise” scores (panel b) for all listeners ( $n = 64$ ) in three End Pitch conditions in the Peak Height-End Pitch (HT) stimulus set



The effect of the two-way interaction of Language by End Pitch in the perception of emphasis is shown in Figure 17a. Clearly, British English listeners rated Level 1 differently from Levels 2 and 3: the difference between L% and H% discussed above is clearly reflected in their scores. By contrast, Dutch listeners did not exhibit any such sensitivity to the difference between L% and H%. Figure 17b shows the interaction of Language by End Pitch for the perceived surprise in questions: even though both groups of listeners perceived little meaning difference between Level 2 and Level 3, British English listeners perceived an observably larger meaning difference between Level 1 and Level 2 than Dutch listeners.<sup>13</sup> We therefore would appear to have a case

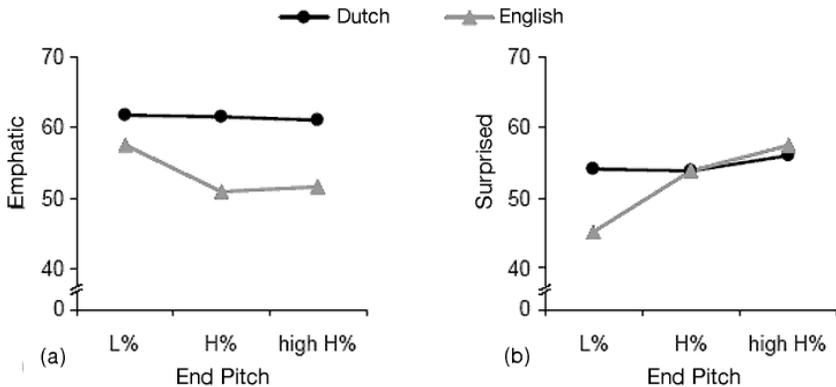
<sup>12</sup> The effect of End Pitch on the perception of “surprised” found in the AT stimulus set is very similar, though listeners made a smaller meaning distinction between Level 1, which scored 44.1, and Level 3, which scored 52.5, than in the HT stimulus set.

<sup>13</sup> The ratings of Dutch listeners are identical to those of English listeners as regards the statements. The interaction of End Pitch by Language is thus not shown for this sentence type.

in which one language acknowledges meaning differences between L% and H%, while the other language does not, or does so to a lesser extent. This can be considered a Type 3 difference (Fig. 1c). The relative insensitivity of Dutch listeners to the difference between L% and H% is very interesting but clearly requires more research, since our findings present only a partial picture obtained from three end pitch heights and find no corroboration in any earlier research findings we are aware of.

**Figure 17**

Mean “emphasis” scores in three End Pitch conditions (panel a) and mean “surprise” scores in three End Pitch conditions in questions (panel b) for Dutch ( $n = 25$ ) and English ( $n = 25$ ) listeners in the Peak Height-End Pitch (HT) stimulus set



## 5 Summary and Conclusions

The work reported here concerned an investigation of the way in which two languages implement postulated universal paralinguistic form-function relations embodied in two biological codes, Ohala's Frequency Code (Ohala, 1983, 1984) and Gussenhoven's Effort Code (2002). Two cross-linguistic perceptual experiments were conducted in British English and Dutch, in which we examined how the two listener groups differed in the perception of “confident” and “friendly” (the Frequency Code, Experiment 1) and “emphatic” and “surprised” (the Effort Code, Experiment 2). A priori, languages can differ in the perception of paralinguistic intonational meaning in a number of ways. We hypothesized that the difference between British English and Dutch would be a Type 1 difference, that is, a difference in the degree to which pitch differences trigger meaning differences. This Hypothesis was based on the existence of a difference in standard pitch range between the two languages (de Bot, 1982; de Pijper, 1983; Willems, 1982), since standard pitch range would appear to affect the perception of intonational meaning (Rietveld et al., 1999). Motivated by Rietveld et al.'s finding, we proposed the Relative Scale to picture how the difference in the standard pitch range between British English and Dutch would lead to a Type 1 difference. According to the Relative Scale, speakers project their standard pitch range onto a given semantic scale in a relative fashion, such that speakers of the narrow-range language, Dutch,

associate a larger meaning difference with a given interval of pitch variation than speakers of the wide-range language, British English (Fig. 2b).

Our Hypothesis of Type 1 difference was borne out for the perception of “friendly” as signaled by pitch register (Fig. 8), “confident” as signaled by pitch register (Fig. 6), “emphatic” as signaled by peak height (Fig. 10), and “surprised” as signaled both peak height (Fig. 12) and pitch register (Fig. 13). Except in the perception of “friendly,” the Type 1 differences can be explained by the difference in standard pitch range, as pictured in the Relative Scale.

Moreover, in addition to Type 1 differences, we unexpectedly found a Type 2 difference in the perception of “emphatic” as signaled by pitch register variation (Fig. 11). A Type 2 difference between two languages means that they display opposite relations between perceived meaning and pitch variation. Arguably, we also found a Type 3 difference, which was defined to mean that a correlation between pitch and perceived meaning is present in one language but absent in the other. This occurred in the perception of “surprised” as signaled by peak alignment (Fig. 14) and end pitch (Fig. 17b), as well as in the perception of “emphatic” as signaled by end pitch (Fig. 17a).

In search of an explanation for these unexpected findings, we arrived at the conclusion that there are two further aspects to be considered in our account of language-specificity in paralinguistic intonational meaning. One is that languages may make different choices whenever conflicting meanings are derived from the biological codes. Such conflicts arise naturally in a situation where three biological codes employ a single phonetic parameter, F0. In particular, pitch register would appear to be used to signal friendliness in British English on the basis of the Frequency Code, but emphasis in Dutch on the basis of the Effort Code. When pitch register was raised, British English listeners perceived a lower degree of emphasis, as a result of perceiving a higher degree of friendliness, while Dutch listeners perceived a higher degree of emphasis. The fact that British English listeners made a larger meaning distinction than Dutch listeners between the lowest and the highest pitch registers in the perception of friendliness is consistent with this explanation. The association of pitch register with emphasis precludes the use of pitch register for the expression of friendliness, or at least makes such use less effective.

The Type 3 differences we uncovered amount to the finding that Dutch speakers fail to exploit the phonetic space of peak alignment (Fig. 14) and end pitch (Fig. 17) to anything like the same degree as British English listeners. A gradient increase in the perceived surprise was present across all three levels of Peak Alignment in British English, but was only present from Level 1 to Level 2 in Dutch. This can be attributed to the fact that variations in tonal alignments are used to signal “routineness” in British English (Gussenhoven, 1984). As for End Pitch, British English listeners appeared to associate L% with a higher degree of emphasis and H% a higher degree of surprise, while Dutch listeners did not make any such distinction. Apparently, speech communities may attribute more or less importance to pitch variation at the phrase end. We do not have sufficient data to be able to speculate on the reason why Dutch listeners chose to ignore this variation.

It should be stressed that the differences between Dutch and English listeners found here cannot be due to the fact that they rated different stimulus sets. In a separate study (Chen, 2005) on the perception of paralinguistic intonational meaning in a second language (L2), the author asked Dutch learners of British English to rate the English stimuli and British English learners of Dutch to rate the Dutch stimuli used in the present study. Systematic differences in the ratings were found between L2 learners and their native-speaker counterparts from this study, who judged the same stimulus set.

An important conclusion to be drawn from our investigation is that in spite of their close genetic relation and similarities in their intonation systems, British English and Dutch differ significantly in the implementation of the biological codes. The presence of a Type 2 difference and Type 3 differences in particular calls the view into question that paralinguistic intonational meaning is universal. While our findings do not speak against the universal existence of the biological codes, they do suggest that these codes are implemented in language-specific ways.

The findings suggest a number of topics for further research. First, this research was conducted in British English and Dutch. An obvious next step is to extend the present investigation to other language groups and shed light on the perception of paralinguistic intonational meaning in a wider variety of languages. Second, this study made use of stimuli in a female voice. However, there are significant gender differences in the use of pitch range for communicative purposes. For example, according to van Leeuwen (1999, p. 109), “where men raise their voice to a higher pitch, women may do the reverse—for instance in newsreading.” Specifically, male newsreaders tend to speak at a higher pitch register (and with a higher intensity as well) when on air than in ordinary daily speech, whereas female newsreaders use a lower pitch register. This difference may well lead to different results in the perception of confidence (Experiment 1) and emphasis (Experiment 2) as signaled by pitch register if stimuli in a male voice are used. It will therefore be interesting to examine the perception of paralinguistic intonational meaning as signaled by a male voice.

*Received: January 15, 2004; first revision received: July 30, 2004;  
second revision received: February 1, 2005; accepted: February 3, 2005*

## References

- ANDERSON, N. H. (1961). Scales and statistics: Parametric and nonparametric. *Psychological Bulletin*, **58**, 305–306.
- BOERSMA, P., & WEENINK, D. D. (1996). PRAAT 3.4: a system for doing phonetics by computer. Institute of Phonetics Sciences of the University of Amsterdam, Report 132. An updated copy of this manual available at <<http://www.praat.org>>.
- BRAUN, B. (2004). *Production and perception of contrastive and non-contrastive themes in German*. Ph.D. dissertation, University of Saarland.
- CHEN, A. J. (2005). *Universal and language-specific perception of paralinguistic intonational meaning*. Ph.D. dissertation. Radboud University Nijmegen. Utrecht: LOT.
- COLLIER, R., & HART, J. (1981). *Cursus Nederlandse Intonatie*. Louvain: Acco/De Horstink.
- CRUTTENDEN, A. (1997). *Intonation* (2nd ed.). Cambridge: Cambridge University Press.

- de BOT, K. (1982). *Visuele feedback van intonatie*. Ph.D. dissertation. Radboud University Nijmegen.
- de JONG, K. J. (1995). The supraglottal articulation of prominence in English: Linguistic stress as localized hyperarticulation. *Journal of the Acoustical Society of America*, **97**, 491–504.
- de PIJPER, J. (1983). *Modeling British English Intonation*. Dordrecht, the Netherlands: Foris.
- DEXTER, F. & CHESTNUT, D. H. (1995). Analysis of statistical tests to compare visual analog scale measurements among groups. *Anesthesiology*, **82**, 896–902.
- GUSSENHOVEN, C. (1984). *On the grammar and semantics of sentence accents*. Dordrecht: Foris.
- GUSSENHOVEN, C. (1999). Discreteness and gradience in intonational contrast. *Language and Speech*, **42**, 281–305.
- GUSSENHOVEN, C. (2002). Intonation and interpretation: phonetics and phonology. In B. Bel & I. Marlien (Eds.), *Proceedings of the speech prosody* (pp. 47–57). Aix-en Provence: Université de Provence.
- GUSSENHOVEN, C. (2004). *The phonology of tone and intonation*. Cambridge, UK: Cambridge University Press.
- GUSSENHOVEN, C. (2005). Transcription of Dutch intonation. In Sun-Ah Jun (Ed.), *Prosodic typology: The phonology of intonation and phrasing* (pp. 118–145). Oxford: Oxford University Press.
- GUSSENHOVEN, C., & CHEN, A. J. (2000). Universal and language-specific effects in the perception of question intonation. In B. Yuan, T. Huang & X. Tang (Eds.), *Proceedings of The Sixth International Conference on Spoken Language Processing*, Vol. I (pp. 91–94). Beijing: China Military Friendship Publish.
- GUSSENHOVEN, C., RIETVELD, T., & TERKEN, J. (1999). Transcription of Dutch intonation: Courseware, version 1.1. <<http://lands.let.kun.nl/todi/todi/home.htm>>.
- HADDING-KOCH, K., & STUDDERT-KENNEDY, M. (1964). An experimental study of some intonational contours. *Phonetica*, **11**, 175–185.
- HARRIS, R. J. (1975). *A primer of multivariate statistics*. London: Academic Press.
- HIRSCHBERG, J., & WARD, G. (1992). The influence of pitch range, duration, amplitude and spectral features on the interpretation of the rise-fall-rise intonation contour in English. *Journal of Phonetics*, **20**, 241–251.
- KRIEMAN, J., GERRATT, B. R., KEMPSTER, G. B., ERMAN, A., & BERKEN, G. S. (1993). Perceptual evaluation of voice quality: review, tutorial, and a framework for future research. *Journal of Speech and Hearing Research*, **36**, 21–40.
- LABOVITZ, S. (1970). The assignment of numbers to rank order categories. *American Sociological Review*, **35**, 515–524.
- LADD, D. R. (1996). *Intonational phonology*. Cambridge: Cambridge University Press.
- MAKAROVA, V. (2000a). Cross-linguistic aspects of intonation perception. In *Proceedings of the 6<sup>th</sup> International Conference on Spoken Language Processing*, vol. III (pp. 452–455). Beijing: China Military Friendship Publish.
- MAKAROVA, V. (2000b). Universal and specific features in intonation perception. In *Proceedings of SICOPS* (pp. 73–81).
- MORTON, E. W. (1977). On the occurrence and significance of motivation-structural rules in some bird and mammal sounds. *American Naturalist*, **111**, 855–869.
- MUNRO, B., & PAGE, E. (1993). *Statistical methods for health care research*. Phil: J. B. Lippincott.
- OHALA, J. J. (1983). Cross-language use of pitch: An ethological view. *Phonetica*, **40**, 1–18.
- OHALA, J. J. (1984). An ethological perspective on common cross-language utilization of F0 of voice. *Phonetica*, **41**, 1–16.
- PATTERSON, D. (2000). *A linguistic approach to pitch range modeling*. Ph.D. dissertation. University of Edinburgh.

- PATTERSON, D., & LADD, D. R. (1999). Pitch range modeling: linguistic dimensions of variation. In J. J. Ohala, Y. Hasegawa, M. Ohala, D. Granville, & A. C. Bailey (Eds.), *Proceedings of the 14<sup>th</sup> International Congress of Phonetic Sciences* (pp. 1169–1172). Berkeley: University of California.
- RIETVELD, T., & GUSSENHOVEN, C. (1995). Aligning pitch targets in speech synthesis: Effects of syllable structure. *Journal of Phonetics*, **23**, 375–385.
- RIETVELD, T., GUSSENHOVEN, C., WICHMANN, A., & GRABE, E. (1999). Communicative effects of rising and falling pitch accents in British English and Dutch. In *Proceedings of the ESCA workshop on dialog and prosody* (pp. 111–116).
- RIETVELD, T., & van HEUVEN, V. J. (1997). *Algemene fonetiek*. Bussum: Uitgeverij Coutinho b.v.
- RIETVELD, T., & VERMILLION, P. (2003). Cues for perceived pitch register. *Phonetica*, **60**, 261–272.
- SCHERER, K. R. (2000). A cross-cultural investigation of emotion inferences from voice and speech: implications for speech technology. In B. Yuan, T. Huang And X. Tang (Eds.), *Proceedings of the Sixth International Conference on Spoken Language Processing*, Vol. 2 (pp. 379–382). Beijing: China Military Friendship Publish.
- STEVENS, S. S. (1946). On the theory of scales of measurement. *Science*, **103**, 677–680.
- SWERTS, M., BOUWHUIS, D. G., & COLLIER R. (1996). Melodic cues to the perceived “finality” of utterances. *Journal of the Acoustical Society of America*, **96**(4), 2064–2075.
- ’t HART, J. (1998). Intonation in Dutch. In D. Hirst & A. Di Cristo (Eds), *Intonation systems*, (pp. 96–111). Cambridge: Cambridge University Press.
- ’t HART, J., & COHEN, A. (1973). Intonation by rule: a perceptual quest. *Journal of Phonetics*, **1**, 309–327.
- ’t HART, J., COLLIER, R., & COHEN, A. (1990). *A perceptual study of intonation*. Cambridge: Cambridge University Press.
- van BEZOOIJEN, R. (1984). *Characteristics and recognizability of vocal expressions of emotion*. Dordrecht: Foris.
- van LEEUWEN, T. (1999). *Speech, music, sound*. London: Macmillan.
- WARD, G., & HIRSCHBERG, J. (1985). Implicating uncertainty. *Language*, **61**, 747–776.
- WEWERS, M. E., & LOWE, N. K. (1990) A critical review of visual analog scales in the measurement of clinical phenomena. *Research in Nursing and Health*, **13**, 227–236.
- WILLEMS, N. (1982). *English intonation from a Dutch point of view: An experimental-Phonetic investigation of English intonation produced by Dutch native speakers*. H. I. Ambacht: Intercontinental Graphics.

## **Appendix 1**

Source utterances for the experimental stimuli in English and Dutch in Experiment 1. They are grouped by speech act in each language. The accented syllable in each sentence is in capitals.

### **English**

#### Information

1. How many DOLLars should I change?
2. How many PHOTocopies should I make?
3. Who will give the presenTATion?
4. What's the LLevel of this course?

#### Request

1. Would you mind turning off the RADio?
2. Could you please send us your appliCATION?
3. Could you please carry that TABLE down the stairs?
4. Could you please move your LUGgage out of the way?

#### Instruction

1. You must warn the DAYnurse.
2. You should contact your PArents.
3. You should fill out the CLAIM form.
4. Please send us the MOney order.

### **Dutch**

#### Information

1. Hoe veel DOLLars moet ik wisselen?
2. Hoe veel koPIEN moet ik maken?
3. Wie gaat de presenTATie geven?
4. Wat is het niVEAU van deze cursus?

#### Request

1. Zou je de RADio uit willen zetten?
2. Zou je ons je AANvraag willen opsturen?
3. Zou je de TAFel even naar beneden kunnen dragen?
4. Zou je je baGAge even kunnen wegzetten?

#### Instruction

1. Je moet de DAGzuster waarschuwen.
2. Je moet contact opnemen met je OUders.
3. Je moet het declaRATie formulier invullen.
4. Stuur ons de beTAALopdracht.

## Appendix 2

Source utterances for the fillers in English and Dutch in Experiment 1. They were grouped by speech act in each language. The accented syllable in each sentence is in capitals. Utterances that were also used as source utterances for the experimental stimuli are in italics.

### English

#### Information

1. How many STUdents will we get?
2. What are the reQUIREments for this job?
3. *What's the LLevel of this course?*

#### Request

1. Could you please cancel my aPPOINTment?
2. Would you mind clearing your TABLE?
3. *Could you please carry your LUuggage out of the way?*

#### Instruction

1. You must send us your PHOTograph.
2. Please enter your PASSword.
3. *Please send us the MOney order.*

### Dutch

#### Information

1. Hoe veel STUdenten krijgen we?
2. Wat zijn de verEISten voor die baan?
3. *Wat is het niVEAU van deze cursus?*

#### Request

1. Zou u mijn AFspraak willen afzeggen?
2. Zou u uw TAFel willen afruimen?
3. *Zou je je baGAge even kunnen wegzetten?*

#### Instruction

1. U moet ons uw FOto sturen.
2. Geef alstublieft uw WACHTwoord.
3. *Stuur onze de beTAALopdracht.*

## Appendix 3

Source utterances for the experimental stimuli in English and Dutch in Experiment 2. They are grouped by sentence mode in each language. The accented syllable in each sentence is in capitals.

### **English**

#### Statements

1. You asked her for a JOB interview.
2. They will beLIEVE the man.
3. They were doing the GARdening.

#### Questions

4. Did you ask her for a JOB interview?
5. Will they beLIEVE the man?
6. Were they doing the GARdening?

### **Dutch**

#### Statements

1. Je hebt haar naar een BAAN gevraagd.
2. Ze zullen de man geLOven.
3. Ze waren de TUIN aan 't doen.

#### Questions

4. Heb je haar naar een BAAN gevraagd?
5. Zullen ze de man geLOven?
6. Waren ze de TUIN aan 't doen?

## Appendix 4

Source utterances for the practice stimuli and end-of-the-list stimuli in English and Dutch in Experiment 2. They are grouped by sentence mode in each language. The accented syllable in each sentence is in capitals.

### **English**

#### Statements

1. They doubted our HOnesty.
2. You are going to TELL her about them.
3. There are MORE like these.
4. We will see the Animals.
5. You are going to neGOtiate with them.

#### Questions

6. Did they doubt our HOnesty?
7. Are you going to TELL her about them?
8. Are there MORE like these?
9. Will we see the Animals?
10. Are you going to neGOtiate with them?

### **Dutch**

#### Statements

1. Ze hadden twijfels over onze EERlijkheid.
2. Je gaat haar het verTEllen.
3. D'r zijn er nog MEER van
4. We zullen de DIEnen nog zien.
5. Je gaat met ze onderHANdelen.

#### Questions

6. Hadden ze twijfels over onze EERlijkheid?
  7. Ga je het haar verTEllen?
  8. Zijn er nog MEER van?
  9. Zullen we de DIEnen nog zien?
  10. Ga je met ze onderHANdelen?
-