

Aoju Chen
Max Planck Institute for Psycholinguistics
aoju.chen@mpi.nl

Abstract

This paper addressed the question of how British English, German and Dutch listeners differ in their perception of continuation intonation both at the phonological level (Experiment 1) and at the level of phonetic implementation (Experiment 2). In Experiment 1, preference scores of pitch contours to signal continuation at the clause-boundary were obtained from these listener groups. It was found that among contours with H%, British English listeners had a strong preference for H*L H%, as predicted. Unexpectedly, British English listeners rated H* H% noticeably more favourably than L*H H%; Dutch listeners largely rated H* H% more favourably than H*L H% and L*H H%; German listeners rated these contours similarly and seemed to have a slight preference for H*L H%. In Experiment 2, the degree to which a final rise was perceived to express continuation was established for each listener group in a made-up language. It was found that although all listener groups associated a higher end pitch with a higher degree of continuation likelihood, the perceived meaning difference for a given interval of end pitch heights varied with the contour shape of the utterance final syllable. When it was comparable to H* H%, British English and Dutch listeners perceived a larger meaning difference than German listeners; when it was comparable to H*L H%, British English listeners perceived a larger difference than German and Dutch listeners. This shows that language-specificity in continuation intonation at the phonological level affects the perception of continuation intonation at the phonetic level.

Keywords: continuation intonation, language-specificity, perception, phonology, phonetic implementation.

Languages: Bengali, Bulgarian, Dutch, British/American English, French, German, Greek, Hungarian, Italian, Japanese, Norwegian, Mandarin Chinese, Miao, Spanish.

Language-specificity in the perception of continuation intonation

1. Introduction

It has been observed for many languages that high or rising pitch is used at a clause boundary to signal continuation of the clause. Following Quirk, Greenbaum, Leech and Svartvik (1980: 342) and Cruttenden (1997: 69), a clause may be the only unit of a simple sentence, part of a compound sentence, or part of a complex sentence. The sound-meaning relation between the clause-final high or rising pitch and continuation is generally known as the ‘continuation rise’ (e.g., Cruttenden 1997 for British English, Pierrehumbert 1981 for American English, ’t Hart, Collier and Cohen 1990 for Dutch, von Essen 1956 for German, Beckman, Díaz-Campos, McGory and Morgen 2002 for Spanish, Arvaniti and Baltazani for Greek, Vanvik 1979 for East Norwegian, Lahiri and Fitzpatrick-Cole 1999 for Bengali, Avgustinova and Andreeva 1999 for Bulgarian). A recent account of the widespread presence of continuation rise across languages holds that it is the grammaticalisation of the paralinguistic use of high pitch (i.e. the use of pitch variations to signal different degrees of a certain meaning) stemming from the Production Code, as argued by Gussenhoven (2002).¹ On the assumption that there is a correlation between utterances and exhalation phases, the Production Code associates high pitch with the beginning of an utterance and low pitch with the end of an utterance. Interpretation of the Production Code includes that high beginnings signal new topics

¹ The Production Code, proposed as part of Gussenhoven's (2002) biologically motivated theory of paralinguistic intonational meaning, is the communicative exploitation of the physiological condition that the generation of subglottal air pressure responsible for the vibration of the vocal cords is tied to the exhalation phase of the breathing process and there is a fall-off of subglottal air pressure towards the end of the exhalation phase.

and low beginnings continuation of topics. At utterance end, high pitch signals continuation, low pitch finality (Gussenhoven 2002: 51).

Quite generally, grammaticalisation of paralinguistic usage of pitch can differ from language to language. For instance, West Germanic languages use pitch accents to mark focused parts of sentences, whereas Japanese suspends downstep in focused sentence constituents (Pierrehumbert and Beckman 1988, Gussenhoven 2002: 51–52). Equally, languages with the same grammaticalisation may differ in the phonetic implementation of the intonational morphemes involved. For example, H*L L% is the typical contour used in declaratives in many Germanic languages and declaratives with narrow focus in Romance languages, but these languages can differ in the temporal alignment of the pitch peak and consequently the temporal alignment of the following fall.² For example, in English and German, the peak tends to occur at or near the end of the stressed syllable and the fall starts between the stressed and the following syllables, whereas in Italian, the peak occurs early in the stressed syllable and the fall begins before the following syllable (e.g., Ladd 1996: 128). It is also known that languages can have more than one grammaticalised form for a given meaning. A case in point is Dutch, where four phonologically different contours may occur in questions (i.e. wh-questions, yes-no questions and declarative questions), H*L H%, H* H%, L*H H%, and L* H% (Haan 2001: 111–113). In this case, languages can differ in their preferred pitch contour in different types of questions.

When languages have more than one grammatical form for a given communicative function, the preferred contour may be taken to be the contour that is most frequently used or is perceived to be most natural sounding by native speakers of a

² Pitch contours are transcribed following the ToDI transcription system (Gussenhoven, Rietveld and Terken 1999, Gussenhoven 2004)

language. In West Germanic, there are two grammaticalised forms of the continuation rise, H% and % (i.e. a high boundary tone and the absence of a boundary tone), which can in turn be preceded by different pitch accents, including H*, H*L, and L*H. As a result, there is more than one pitch contour that can be used on a non-final clause to signal continuation (hereafter continuation contour). The first goal of the present study is to examine language-specific differences in the preferred continuation contour in English, German and Dutch, from the perspective of perception.

1.1. Hypothesis 1: preferred continuation pitch contour

Delattre (1965) is the first one to note a difference in continuation contour between English and German, although he was not particularly concerned with signalling of continuation at a clause boundary. In spoken corpora made up of at least five minutes of spontaneous speech by educated native speakers of German, French, Spanish and American English, Delattre observed that continuation was signalled by a contour with a predominant ‘falling portion’ followed by a relatively weak rise in English, but by a contour with a distinctive rising shape in German, French and Spanish.³ This difference between English and German was confirmed by Grover, Jamieson and Dobrovolsky (1987) in their study on the role of intonation in foreign accent. In order to gain insight of how continuation was signalled in their L2 learners’ native languages (i.e. English, German and French), Grover et al. measured the ‘slope’ (Delattre 1963, as cited in Grover et al. 1987) covering the stressed syllable of the word before the conjunct in each of the sentences containing coordinated phrases/clauses read by native speakers of

³ Delattre seemed to ignore the rising portion in English in his conclusion and contended that continuation was signalled by a falling contour in English but a rise contour in the other three languages. Grover, Jamieson and Dobrovolsky (1987) followed Delattre in this overstatement.

the three languages. The slope was obtained by dividing the maximum F_0 change ($=F_0$ at a later time point – F_0 at an earlier time point) by the time over which the change took place. It was found that by and large the slopes were positive in German and French utterances but negative in English utterances, indicating that there was a pitch rise over the stressed syllable preceding the conjunct in German and French but a pitch fall in English. Note that in Grover et al.'s measurement, the final rising portion of the pitch contour in English could not have affected the slope value because the minimum F_0 was achieved before the final rise.

In the autosegmental-metrical framework, Delattre's and Grover et al.'s findings may be interpreted to mean that English, German, French and Spanish employ a non-low final boundary tone to signal continuation but differ in the nuclear pitch accent. That is, English would seem to favour the fall-rise contour ($H^*L H\%$), but German (as well as French and Spanish) a rise contour, which may be represented in different phonological forms, such as $L^*H H\%$, $H^* H\%$, $L^*H \%$, or $H^* \%$. This interpretation is fed with empirical evidence by Sanders (1996) and is to some extent shared by Féry (1993) regarding English and German respectively. In a study of boundary marking in British English, Sanders (1996: 102–105) found that at the boundary between two coordinated clauses and/or between a main clause and a subordinate clause, native speakers of British English rated contours with a final rise as more natural-sounding than contours with a final fall. These contours included the 'half rise' contour, in which the nuclear fall was followed by a rise to a default mid level; the 'full rise' contour, in which the nuclear fall was followed by a rise to a point as high as the pitch peak of the nuclear fall; and a 'virtual rise' contour, in which the nuclear fall was not followed by a rise but by a pitch point realised at the default mid level. In her description of German

intonational system based on a corpus of 100 sentences read by three native speakers of German, Féry (1993: 85-89) noted that the rising pitch accent L*H was typically used to signal the speaker's intention to continue, in addition to his invitation for an answer or a confirmation from the hearer. When followed by H%, it was frequently used in tag words (e.g. *ja* meaning 'go on', *nicht and nicht wahr* meaning 'isn't it'); when followed by % (also referred to as the progradient intonation, following von Essen 1956), it was commonly found at the end of an Intonational Phrase (IP) of a sentence.

In connection with such a difference between two genetically closely related languages, English and German, the interesting question arises as to what would be the preferred continuation contour in Dutch, a language that not only resembles the intonation of British English in numerous ways, but also is at the same time believed to sound rather similar to German. In a production study, 't Hart and Cohen (1973, as cited in 't Hart, Collier and Cohen 1990: 101–102) asked six native speakers of Dutch to read a series of Dutch proverbs, fourteen of which contained a clause boundary, and found that in two-thirds of the renditions, the clause boundary was marked by means of one of the three pitch contours, A2 (H*L H%), 1 (H* %) and E (!H*L). The distribution of these three contours in these instances was as follows: over 50% (H*L H%), about 30% (H* %), and about 20% (!H*L). This distribution was largely confirmed in the same study by a sample of readings of sentences taken from a connected prose passage, though contour !H*L was found to be used less frequently (in only 8% of the instances), and contour H* % more frequently (in about 42% of the instances). It may thus be inferred that H*L H% and H* % typically serve as the continuation contours. On the other hand, it is suggested (Gussenhoven et al. 1999) that H* % and L*H % are used to

signal continuation utterance-internally, whereas H*L H% is used to signal a question, a reminder or a suggestion.

In view of the findings and observations in earlier studies considered above, we predicted that (1) British English listeners prefer H*L H%; (2) German listeners prefer L*H %; and (3) Dutch listeners may not have a clear preference between the fall-rise contour and the rise contours. To keep the present study to a controllable length, we focused on language-specific preference among pitch contours with a complete final rise or a high boundary tone, i.e. H*L H%, L*H H%, and H* H%. We assumed the following hypothesis on the preferred continuation contour(s) in each of the three languages:

Hypothesis 1:

- (1) H*L H% is preferred to L*H H% and H* H% in British English;
- (2) (in the absence of L*H %) L*H H% is preferred to H*L H% and H* H% in German;
- (3) H*L H%, H* H% and L*H H% are similarly favoured in Dutch.

1.2. Hypothesis 2: perception of final rise as a continuation cue

According to Gussenhoven (2002), speakers/listeners, regardless of language background, can exploit the phonetic implementation of the final rise to signal/interpret the degree of continuation of an utterance; the higher the final rise ends, the more non-final the utterance sounds. Previous studies have shown that differences in intonational grammar across languages can lead to language-specificity in the perception of paralinguistic intonational meaning. In Chen, Gussenhoven and Rietveld (2004), native speakers of Dutch and British English listened to utterances varying in pitch-range

related parameters in their native language and judged each utterance on a number of speaker attributes including 'surprised'. It was found that British English listeners associated a later peak with a higher degree of surprise, whereas Dutch listeners hardly perceived any meaning differences when peak alignment was varied. The authors accounted for this difference between British English and Dutch listeners by the fact that peak delay is exploited to signal nonroutiness in British English (Gussenhoven 1984: 217–220), but is an uncommon phenomenon in the Dutch intonational system. Interestingly, language-specific perception of paralinguistic intonational meaning as triggered by differences between their intonational grammars also occur when listeners are presented with an unknown language. In a study on the perception of question intonation in a made-up language, Gussenhoven and Chen (2000) found that Hungarian, Mandarin Chinese and Dutch listeners all perceived utterances with a higher peak more frequently as questions. However, Hungarian listeners, who were accustomed to the use of peak height to signal questions in their native language, perceived a significantly larger meaning difference for a given interval of peak heights than Dutch and Mandarin Chinese listeners, who used peak height to a lesser extent (in Dutch) or not at all (in Mandarin Chinese) for the signalling of questions. These findings led to the question of how language-specificity in the preferred continuation contour would affect the perception of final rise as a cue for continuation. The second goal of this study is thus to shed light on this question in the context of British English, German and Dutch.

On the basis of the above findings, it may be expected that listeners are most sensitive to variations in the final rising portion of their preferred continuation contour and hence the corresponding meaning differences. When asked to judge how likely an utterance, which is syntactically complete but semantically ambiguous between being

complete and being incomplete, is to be continued by another utterance, British English, German and Dutch listeners will differ in how much difference in continuation likelihood they perceive for a given interval of end pitch values, depending on the pitch contour of the utterance. Specifically, we proposed the following hypothesis:

Hypothesis 2:

- (1) When the final rise is part of a rise contour, German listeners will perceive a larger difference in continuation likelihood for a given interval of final rises than Dutch listeners and British English listeners;
- (2) when the final rise is part of a fall-rise contour, British English listeners will perceive a larger difference in continuation likelihood than Dutch and German listeners.

Two experiments were designed to test the two hypotheses separately. Section 2 reports Experiment 1 testing Hypothesis 1 and section 3 reports Experiment 2 testing Hypothesis 2.

2. Experiment 1: testing Hypothesis 1

A conventional way to obtain listeners' preference for one particular contour at a clause boundary is to ask them to listen to sentences with two clauses, the first of which is realised with different pitch contours, and judge for each sentence how the connection is made between the two clauses in terms of intonation, for example, on a five-point scale. In this scale, '1' and '5' represent 'hardly appropriate' and 'highly appropriate' respectively, whereas '2', '3', and '4' represent intermediate positions. The contour that has the highest mean score across different sentences will be considered the preferred

one. However, an obvious disadvantage of this method is that a listener may independently assign the same score to two different contours but still have a slight preference for one, as pointed out by Scheffé (1952). We therefore rejected this method and adopted Scheffé's paired-comparisons paradigm for the present experiment. Intonationally different renditions of a given compound sentence or sentence sequence (i.e. a sequence of two simple sentences) were presented in pairs to listeners. Their task was to judge which of the two renditions sounded better in terms of how the two clauses in each rendition were intonationally connected, and indicate the degree to which this was the case on a seven-point scale (-3, -2, -1, 0, 1, 2, 3). Scores -1, -2, and -3 meant an increasing preference for the first rendition, and scores 1, 2, and 3 an increasing preference for the second rendition.

2.1. Stimuli

Four compound sentences (1a, 2a, 3a, 4a) and four sentence sequences (1b, 2b, 3b, 4b) (see Appendix 1) were drawn up in each of the three languages, exemplifying two types of continuation, i.e. sentence-nonfinal continuation and sentence-final continuation respectively. The compound sentences were limited to sentences containing two clauses coordinated by either 'and' or 'but'. Together with the four sentence sequences, they served as the 'source expressions' from which the stimuli were generated. The source expressions were comparable across languages in terms of lexical material, semantic content, word order, and sentence structure.

In one half of the source expressions in each language, the accented word of the first clause contained one single syllable with a voiced coda; in the other half of the source expressions, the accented word of the first clause contained two or more

syllables with the stressed syllable followed by at least another syllable. In this way, the source expressions exemplified two pitch accent positions, IP-final vs. non IP-final (abbreviated final vs. non-final). Pitch-accent position was taken into account because it may affect choice of pitch contour. For example, Fery (1993: 91) suggested that it was 'slightly marked' to have H*L H% on IP-final position in Standard German. Similarly, Grabe (1998: 138) observed that H*L H% did not occur on IP-final position in her Standard German read-speech corpus.

In addition, to maximise the effects of continuation type on choice of pitch contour, the first clauses of the four compound sentences and those of the four sentence sequences were lexically identical such that they formed four minimal pairs. Examples of such a minimal pair are given in English, German and Dutch in (A), (B) and (C), in which the accented syllables are in capitals. The distribution of continuation types and pitch accent positions over the eight source expressions is displayed in Table 1.

(A) a. The story is too LONG. The plot is boring.

b. The story is too LONG but is fun to read.

(B) a. Die Geschichte ist zu LANG. Der Inhalt ist langweilig.

b. Die Geschichte ist zu LANG, aber angenehm zu lesen.

(C) a. Het verhaal is te LANG. De plot is saai.

b. Het verhaal is te LANG maar is leuk om te lezen.

Table 1. *The distribution of continuation types and pitch accent positions over the eight source expressions. Source expressions coded with the same letter have the same type of continuation. IP stands for Intonational Phrase.*

Continuation type	Sentence-final		Sentence-nonfinal	
	IP-final	IP-nonfinal	IP-final	IP-nonfinal
Source expressions	1a, 2a	3a, 4a	1b, 2b	3b, 4b

Each of the three pitch contours, H* H%, H*L H% and L*H H%, was imposed on the first clause of each source expression. The onset was fixed as %L. Because in natural speech the final rising portion of a given rise contour can be realised differently, two realisations of the final rising portion were included for each of the three pitch contours (see Fig. 1). This gave us two phonetic variants for each pitch contour and in total six contour-conditions (3 pitch contours \times 2 variants): H* H%1, H* H%2, H*L H%1, H*L H%2, L*H H%1, and L*H H%2. The second clause of each source expression was assigned the contour %L H*L L%. This gave us six renditions per source expression. For each source expression, pairing the six renditions with each other gave us 15 pairs in the order AB and another 15 pairs in the order BA. We will use the contour-conditions of the first clauses of the two renditions to refer to each pair. For example, the pair H* H%1-H*L H%2 means that the first clause of rendition A has variant 1 of H* H% and the first clause of rendition B has variant 2 of H*L H%. In total, there were 120 contour-condition pairs (15 contour-condition pairs \times 8 source expressions) in the AB order and another 120 in the BA order.

Details on the recordings of the source expressions and speech manipulation in each language are discussed in sections 2.1.1 and 2.1.2 respectively.

2.1.1. Recording

The source expressions were recorded in British English, German and Dutch on DAT tape (48 kHz in 16 bits) in the sound-attenuated studio of the Faculty of Arts at the Radboud University Nijmegen. They were read by a phonetically trained male native-speaker of Dutch, whose pronunciation of Standard Southern British English and Standard German is also native-like, as judged by native speakers of British English and German with phonetic training (hereafter expert listeners). The speaker was instructed to read the first clause of each source expression with a fall-rise contour in British English and a rise contour in German and Dutch, and the second clause of each source expression with a fall in all the three languages in a neutral manner. The speaker was also instructed to read the second clause without resetting pitch register (i.e. overall pitch level), when recording the four sentence sequences. Readings of the eight source expressions were digitised at a 32-kHz sampling rate and spliced into separate sound files with one clause per file. Readings of the four ‘first clauses’ and eight ‘second clauses’ with best sound quality (i.e. clear articulation, modest intensity, and no creaky voice) were selected for each language by the experimenter and a trained phonetician. Combining the ‘first clauses’ with their corresponding ‘second clauses’ with a 120-ms pause in between gave us the best recordings of the source expressions.⁴ They were subsequently subjected to speech manipulation.

⁴ According to Sanderman and Collier (1995, as cited in Sanders 1996), contour types have effects on perceived boundary strength only in the absence of pauses. This implies that there is no direct connection between contour types and duration of pauses in predicting boundary strength. It is therefore justified to

2.1.2. Speech manipulation

Speech manipulation was performed by means of the speech processing package Praat (Boersma and Weenink 1996) using the PSOLA technique. Manipulation was conducted mainly to erase the original pitch contours and superimpose new pitch contours, which will be described in detail in the following paragraphs. For two reasons, pre-boundary lengthening manipulation was not applied. First, the domain of lengthening appears to vary from speaker to speaker and differences in duration are small; second, durational differences in the rhymes of the preboundary words spoken in different pitch contours may be hard to detect (Sanders 1996). The pitch manipulation was identical in the three languages for the sake of comparability of the stimuli. A concern may, however, arise that this could lead to a difference in the quality of the stimuli in different languages, considering that languages differ in their phonetic manipulation of pitch contours. To minimise such a quality difference, we examined various sets of pitch values, peak/valley alignments and duration of fall/rise/plateau by checking the intonational acceptability of the resulted speech with both expert listeners, who were familiar with this kind of tasks in British English, German and Dutch, and native speakers of each language. This undertaking resulted in a set of values for pitch height, alignment and duration which appeared to be similarly acceptable in all the three languages.⁵

With respect to manipulation of the four ‘first clauses’, each contour had two phonetic variants, as shown in Fig. 1. Following the expert listeners’ advice, for the

normalise the pause between the two clauses of each source expression. A 120-ms pause appeared to be most appropriate for our stimuli in the three languages.

⁵ The German source expressions were adjusted for duration by lengthening the four ‘first clauses’ by a factor of 1.5, because they were read at a noticeably faster tempo than the eight ‘second clauses’ in German and the British English and Dutch source expressions, and sounded a bit odd to native speakers of German.

sake of naturalness a pre-nuclear high pitch accent (H*L) was superimposed on the stressed syllable of the clause Subject if it was a Noun Phrase, or the stressed syllable of the clause Verb or the copula if the clause Subject was a pronoun. It was realised as a 130-Hz pitch point at the CV boundary of the stressed syllable, 10 ms before the CV boundary if the stressed syllable had a consonant cluster as onset or was preceded by at least another syllable, and 10 ms after the CV boundary if the stressed syllable had a voiced coda or was followed by at least another syllable (Rietveld and Gussenhoven 1995). The contours imposed on the first clauses thus include %L H*L H* H%, %L H*L L*H H%, and %L H*L H*L L%. In the remainder of this text, these contours are referred to as H*L L%, L*H H% and H* H%. The realisation of the nuclear pitch accent in each contour varied depending on pitch accent position. Fig. 1 illustrates the realisation of the nuclear pitch accent on non-final position. H* was realised as a 100-ms rise followed by a high plateau. L*H was realised as a 30-ms low plateau preceded by a 100-ms fall and followed by a 100-ms rise, which was followed by a high plateau. H*L was realised as a 30-ms high plateau preceded by a 100-ms rise and followed by a 100-ms fall, which was followed by a low plateau. The high plateaus of H* H% and L*H H% and the low plateau of H*L were of varying duration and ended at a point that was 100 ms before the pitch point associated with H%. The 100-ms rise of H*, the 30-ms low plateau of L*H, and the 30-ms high plateau of H*L started at the CV boundary of the accented syllable, but 10 ms after the CV boundary when the accented syllable had a voiced coda or followed by another syllable and 10 ms before the CV boundary when the accented syllable was preceded by another syllable (Rietveld and Gussenhoven 1995). On final positions, H* and L*H were realised without the high plateau; H*L was realised without the low plateau.

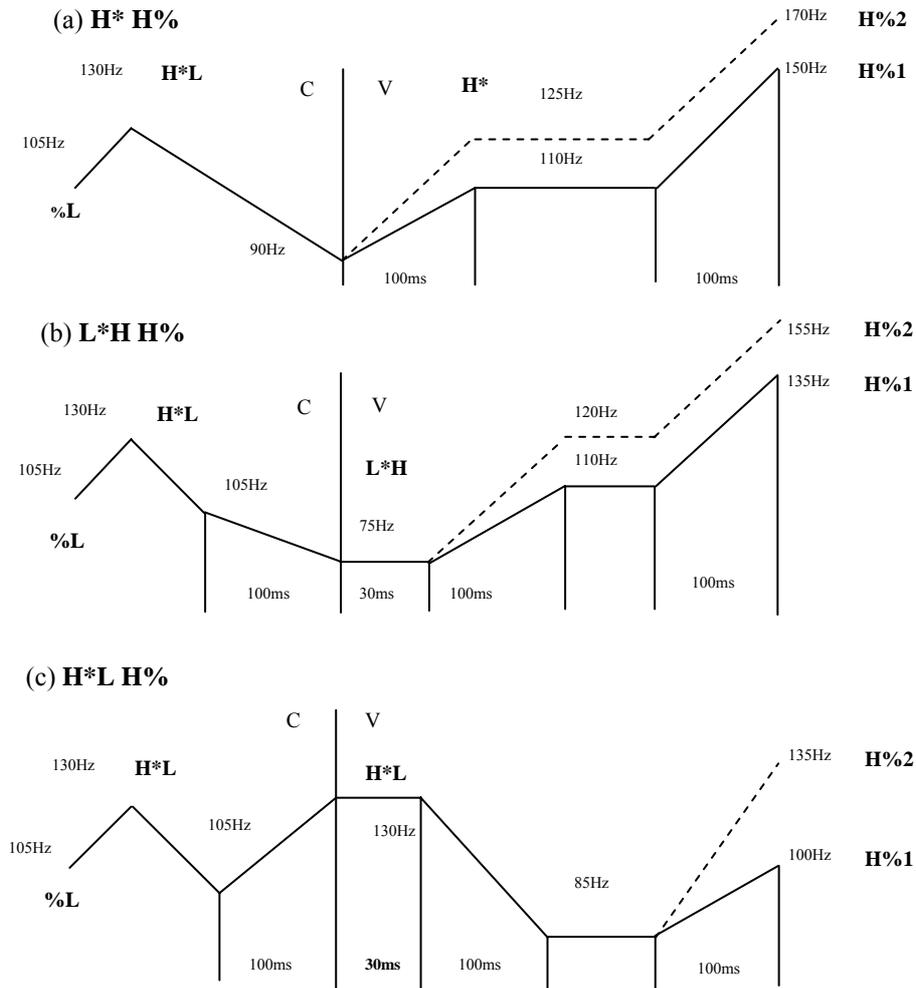


Fig. 1. Schematic representations of %L H*L H* H% (abbreviated H* H%), %L H*L L*H H% (abbreviated L*H H%) and %L H*L H*L H% (abbreviated H*L H%) on non IP-final position with variations in the final rising portion.

The schematic representation of the contour %L H*L H*L L% (on IP-nonfinal position) realised on the eight ‘second clauses’ is shown in Fig. 2. The realisations of pre-nuclear and nuclear pitch accents followed the principles specified for the three continuation contours except that the pitch values were set differently. The placement of prenuclear accent was also somewhat different from that in the first clauses.

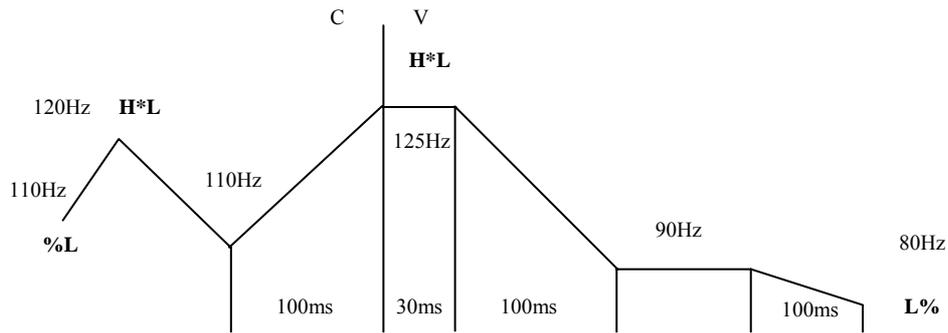


Fig. 2. Schematic representation of the contour %L H* H*L L% (abbreviated H*L L%) in the second clause of each source expression on IP non-final position.

2.2. Test tapes

The 120 stimulus pairs (15 contour-condition pairs \times 8 source expressions) in the AB order in each language were mixed manually and divided into 12 blocks of 10 stimulus pairs. Another four pairs were generated from two utterances comparable to the source expressions in each language. These four pairs were divided into 2 blocks and served as the practice trials. The stimulus pairs and the practice trial pairs were recorded onto DAT tape (48 kHz in 16 bits) with a 1-s pause between the two items in each pair, a 4.5-s pause between pairs, a 7-s pause between blocks, and a 200-ms 300-Hz sine wave preceding each block to signal the beginning of the block. The recording on DAT tape was copied to TDK audio tapes. This gave us test tape AB (approximately 20 minutes) in each language. This procedure was repeated with the order of the two items in each stimulus pair switched from AB to BA but everything else intact. This gave us test tape BA (approximately 20 minutes) in each language.

2.3. Procedure

Thirty-two native speakers of British English (9 men and 23 women), 23 native speakers of German (8 men and 15 women), and 16 native speakers of Dutch (4 men and 12 women) took part in the experiment. Twenty-eight of the English subjects (5 men and 23 women) were undergraduates in the Department of Theoretical and Applied Linguistics at the University of Edinburgh and the other subjects (4 men) were postgraduates from other departments of the same university. German subjects were undergraduates in the Department of Speech Science and Phonetics at the University of Halle (Saale) – Wittenberg. Dutch subjects were recruited from first year students in the Department of English at the Radboud University Nijmegen. All German subjects and twenty-eight of the English subjects participated as part of their course requirements; the other four English and all Dutch subjects were paid a small fee.

Approximately half of the subjects were assigned test tape AB and half test tape BA with stimuli in their native language. All subjects listened to the tape through a high quality recorder/player at an adequate volume in a quiet room. They were instructed by means of written instructions in their native language to pay attention to the intonation of each utterance, judge for each stimulus pair which reading sounded better in terms of how the connection between the two clauses of the expression was made intonationally, and indicate how much better it was on a 7 point scale (-3, -2, -1, 0, 1, 2, 3) printed for each stimulus pair on their score sheets. The actual experiment was preceded by a short practice session (without feedback). After the practice session, subjects were given the opportunity to raise questions.

To control for their language background, German and English subjects were asked to complete a questionnaire at the end of the experiment. As to Dutch subjects, it

was established by means of an informal survey that they were not early bilinguals (i.e. bilinguals who have acquired a second language in their childhood) and mainly came from two neighbouring provinces in the south of the country (i.e. Noord Brabant and Gelderland), whose varieties of Dutch have not been reported to differ intonationally.

2.4. Statistical analyses

2.4.1. Analysis of variance for paired comparisons

Scheffé's (1952) analysis of variance (ANOVA) for paired comparisons was adopted to analyse the present paired comparison experiment. This method was successfully used before in an experiment on the relation between pitch excursion size and perceived prominence (Rietveld and Gussenhoven 1985) and in another experiment on the relation between intonation and perceived speech rate (Rietveld and Gussenhoven 1987). In this analysis, it is assumed that the m variants to be compared can be characterised by parameters $\alpha_1, \alpha_2, \dots, \alpha_m$, and the average preference for variant i over variant j is the difference of the corresponding parameters ($\alpha_i - \alpha_j$). This assumption is referred to as the hypothesis of subtractivity, 'analogous to the interactions in a two-way layout' (of the standard ANOVA). The parameters are considered 'analogous to the main effects in the ... two-way layout' (Scheffé 1952: 386). The significance of the main effects is not tested by the conventional F-ratios but on the basis of estimated scale values of these parameters, $\hat{\alpha}_1, \hat{\alpha}_2, \dots, \hat{\alpha}_m$, and a "yardstick" Y_ϵ . The main effect is significant if the largest and the smallest estimated scale values differ by more than the "yardstick" Y_ϵ . If the main effect is significant, then it can be concluded that overall differences in preference are observed for all variants. Once it is established that there are such overall differences in preference, inferences can be made about whether there is a significant

difference in preference between any two of the variants. If the estimated scale values of variants i and j differ by at least the “yardstick” Y_{ϵ} , it will be concluded that there is a significant difference in preference between i and j , provided that the hypothesis of subtractivity is not violated. If the hypothesis of subtractivity is violated, then $\hat{\alpha}_i - \hat{\alpha}_j$, no longer an unbiased estimate of the mean preference for i over j but still the best unbiased estimate of $\alpha_i - \alpha_j$, ‘measures the relative superiority of i over j in an average sense when i and j are compared with the $m-2$ other variants as well as with each other’ but the former with greater accuracy (Scheffé 1952: 395).

Note that the Scheffé’s ANOVA for paired comparisons assumes that each pair of comparisons is judged once in the order (i, j) and once in the order (j, i) throughout the test.. A related requirement for the experimental design is that the judges who judge the variants in order (i, j) are not the same judges who judge the variants in the order (j, i) ..

2.4.2. Analyses

Frequencies of scores assigned to each contour-condition pair was obtained per source expression for each subject. This gave us a set of analysable data.⁶ Data from 3 German

⁶ In the analyses reported in the earlier version of this text, a data trimming procedure was carried out such that only data from a selected group of subjects (14 English subjects, 10 German subjects and 8 Dutch subjects) were included. This was motivated by the observation that a large number of subjects assigned substantially more positive scores than negative scores, suggesting that they often considered the second utterance in each pair more appropriate than the first utterance. As our hypotheses did not allow us to make accurate predictions on the distribution of negative and positive scores and the ‘second better’ tendency was earlier reported in studies of intonation (e.g., Chorianopoulou 2002, Calhoun 2003) using forced-choice tasks, we felt that to minimize the risk of getting skewed results, it was necessary to exclude data from subjects who assigned a positive score in more than 70% of the cases. The results obtained from the selected set of data turned out to be quite similar to the results obtained from nearly all data, which are reported here. There were two noticeable differences: (1) The preference for H* H% over L*H H% became clearer in English when data from all English subjects were included for analyses; (2) H*L H% was rated slightly more favourably than L*H H% in Dutch when data from all Dutch subjects were included but the opposite held when only a subset of the Dutch data were analysed. For the sake of clarity we do not give the details of the analyses of the subset here, for two reasons. First, since we cannot predict the distribution of scores from the hypotheses, it may not be justifiable to exclude data only under the assumption of a possible strong order effect. Second, the differences between the analyses are marginal. The first difference concerning the results from the English data is a matter of degree. The

subjects could not be used for statistical analyses for the following reasons: (1) from a different age range than the others (1 subject); (2) not a native speaker of German (2 subjects); data from another 3 German subjects who judged the stimuli in the BA order were excluded to make sure that the number of subjects receiving the stimuli in the order BA was the same as those receiving the stimuli in the order AB in the data set. In total, data from 32 English subjects, 16 Dutch subjects and 16 German subjects were included for statistical analyses.

As mentioned in section 2.1, each contour-condition pair was implemented on eight source expressions, exemplifying two types of continuation (i.e. sentence-nonfinal and sentence-final) and two pitch accent positions (i.e. final and non-final). Each contour-condition pair was thus judged eight times by each subject. Since Scheffé's analysis cannot process repeated measures, data of each group of listeners were divided up into eight sets with each set including only data obtained from the lexically identical stimulus pairs. These eight data sets will be referred to by the numbering of the corresponding source expression: set 1a, set 2a, set 3a, set 4a, set 1b, set 2b, set 3b, and set 4b. Twenty-four separate analyses of variance for paired comparisons were then performed on the frequencies of scores per contour-condition pair, one for each data set of each listener group, at the significance level of 0.05.⁷ Main effects of the six contour-conditions and order effects were analysed in each analysis. When necessary, additional statistics were obtained either to find out the generalisability of the findings emerging from the ANOVAs or to further bring out differences between contour-conditions. Details on the additional statistics are given in section 2.5.

second difference, the reversal in the preference for H*L H% and L*H H% in the Dutch data, was very small, as the preference scores for these contours were very similar in both analyses, and does not change the main result that H* H% was rated higher than H*L H% and L*H H%.

⁷ These analyses were conducted by means of a programme written by Toni Rietveld and converted to C by Gies Bouwman from the Radboud University Nijmegen.

2.5. Results and discussion

2.5.1. British English data

Table 2 gives an overview of the estimated scale values per contour-condition and the corresponding “yardstick” $Y_{.05}$ in each data set, in addition to the effect sizes of significant order effects.⁸ The hypothesis of subtractivity was not violated in any of the analyses. The order effects were significant in all analyses but had very small effect sizes ($\eta^2 \leq 0.1$). As the largest and the smallest estimated scale values (in bold in Table 2) in each data set differed by more than the corresponding “yardstick” $Y_{.05}$, it can be concluded that there were significant overall differences in preference among the contour-conditions.⁹

⁸ The measure of effect size used here is η^2 ($= SS_{\text{effect}}/SS_{\text{total}}$).

⁹ Fourteen of the German subjects described their accent as Middle German and two as Berlin German. Fourteen of the English subjects described their accent as Scottish, 12 subjects as Southern and six subjects as Northern. Considering intonational differences among the three varieties of English, one may think that subjects with different dialectal background may not prefer the same contour to signal continuation. We encoded the English data such that each negative score was counted as one preference judgement for the first item in each pair and each positive score one preference judgement for the second item in each pair. This gave us a data set with the frequencies of preference judgements for each of the three contours pooled over all comparing contexts. Subsequently, a mixed-design ANOVA was performed on this data set with frequencies of preference judgements as the dependent variable. It included two within-subject factors, Pitch Contour (3 levels) and Continuation Type (2 levels), and one between-subject factor Variety of English (3 levels). No evidence was found for a difference in preferred continuation pitch contour among the three listener groups. This suggests that listeners speaking different varieties of English judged the stimuli in a similar way. This may be because they were judging stimuli spoken in Southern Standard British English and accordingly switched their perceptual ‘language mode’ to this variety.

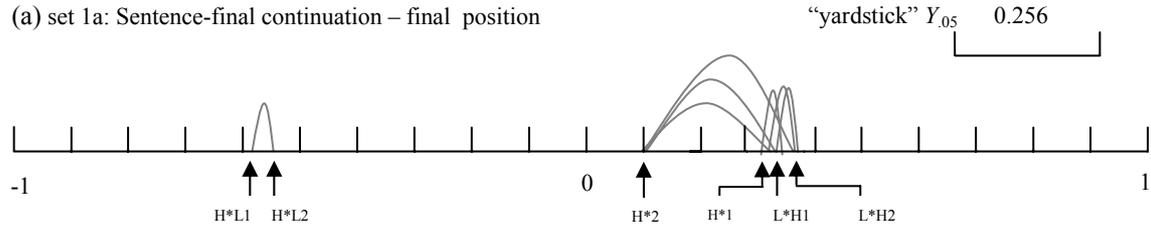
Table 2. *Estimated scale values of the six contour-conditions, “yardsticks” $Y_{.05}$, and effect sizes of significant order effects in each set of English data. The content in each role is indicated in the leftmost column. The largest and the smallest estimated scales values in each data set are marked in bold.*

Continuation type	Final				Non-final			
	Final		Non-final		Final		Non-final	
Pitch accent position	Final	Non-final	Final	Non-final	Final	Non-final	Final	Non-final
Data set	1a	2a	3a	4a	1b	2b	3b	4b
$\hat{\alpha}_{H^*H\%1}$	0.32	0.22	0.09	0.08	0.05	0.22	0.21	0.1
$\hat{\alpha}_{H^*H\%2}$	0.11	0.2	0.16	0.23	0.01	0.08	0.18	0.02
$\hat{\alpha}_{H^*LH\%1}$	-0.58	-0.58	-0.54	-0.64	-0.46	-0.76	-0.73	-0.63
$\hat{\alpha}_{H^*LH\%2}$	-0.55	-0.65	-0.52	-0.59	-0.52	-0.67	-0.6	-0.65
$\hat{\alpha}_{L^*HH\%1}$	0.33	0.38	0.37	0.45	0.54	0.52	0.53	0.55
$\hat{\alpha}_{L^*HH\%2}$	0.36	0.43	0.44	0.46	0.39	0.6	0.41	0.6
“yardstick” $Y_{.05}$	0.26	0.27	0.25	0.25	0.27	0.27	0.26	0.27
Order effects $\delta_{.05}$	*	*	*	*	*	*	*	*
	$\eta^2 = 0.05$	$\eta^2 = 0.09$	$\eta^2 = 0.1$	$\eta^2 = 0.08$	$\eta^2 = 0.07$	$\eta^2 = 0.04$	$\eta^2 = 0.07$	$\eta^2 = 0.07$

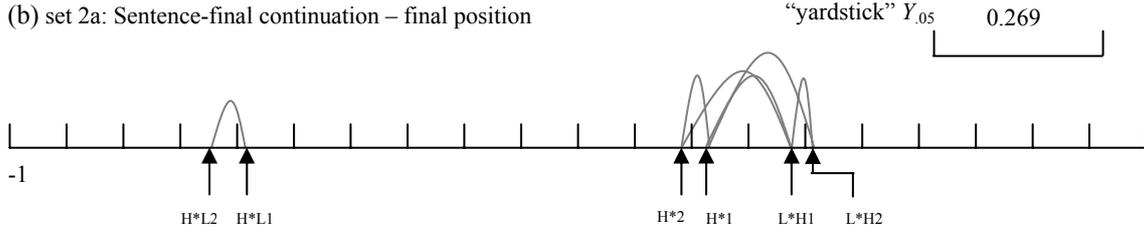
To visualise which contour-conditions differed significantly in preference, the estimated scale value of each contour-condition is plotted per data set in Fig. 3a–h. As all values vary between -1 and 1 , the values are projected on a scale with the two ends anchored by ‘ -1 ’ and ‘ 1 ’. ‘ 0 ’ is indicated on the scale to make it easier to read the values. The six contour-conditions are indicated by pitch accents and variations in the final rising portion (i.e. 1 stands for variation 1 and 2 stands for variation 2, as shown in Fig. 1). In each figure, contour-conditions that did not differ significantly in preference (i.e. $\hat{\alpha}_i - \hat{\alpha}_j - Y_{.05} < 0$) are linked by solid curves. Note that lower scale values correspond to higher preference scores.

Language-specificity in continuation intonation

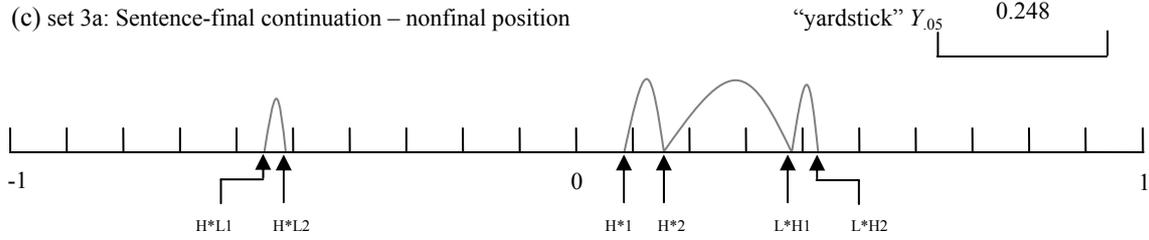
(a) set 1a: Sentence-final continuation – final position



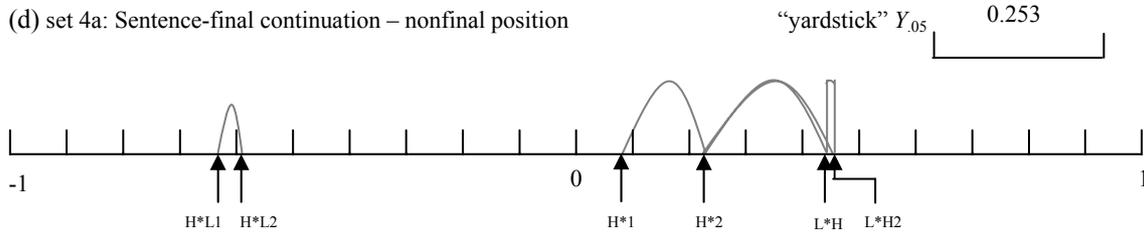
(b) set 2a: Sentence-final continuation – final position



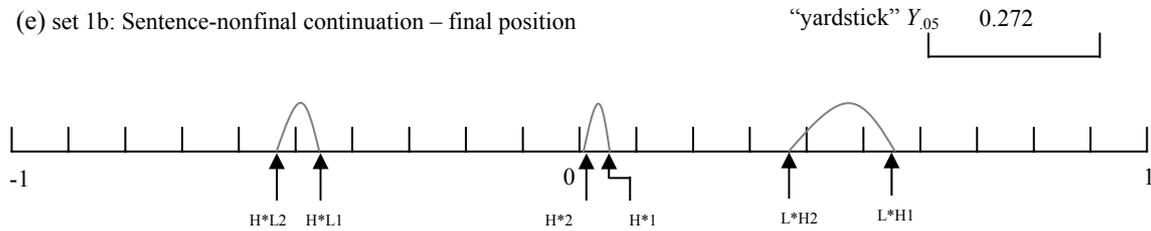
(c) set 3a: Sentence-final continuation – nonfinal position



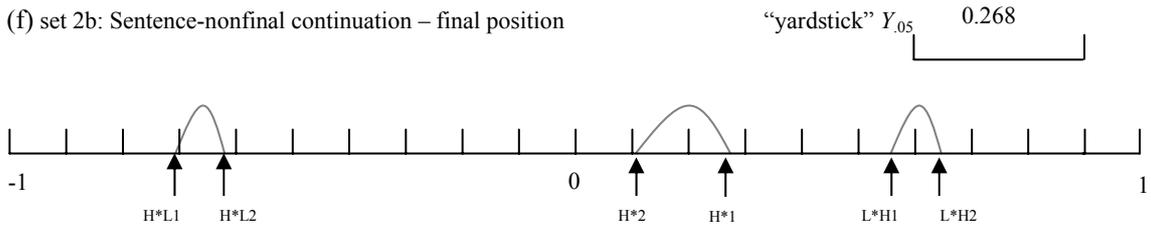
(d) set 4a: Sentence-final continuation – nonfinal position



(e) set 1b: Sentence-nonfinal continuation – final position



(f) set 2b: Sentence-nonfinal continuation – final position



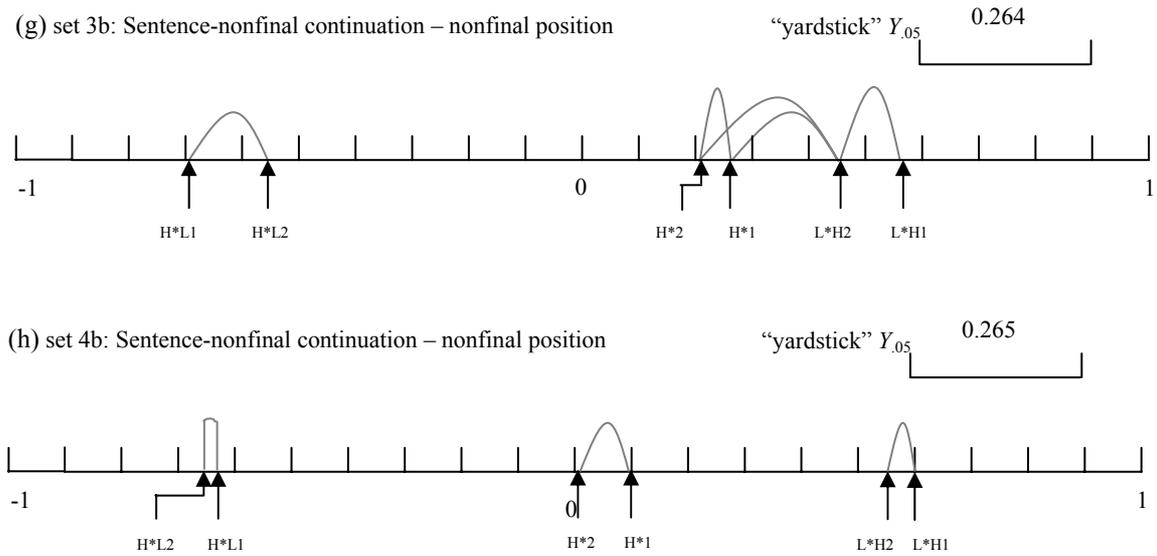


Fig. 3. Estimates of the scale values of the six contour-conditions (H^*1 : $H^* H\%1$, H^*2 : $H^* H\%2$, H^*L1 : $H^*L H\%1$, H^*L2 : $H^*L H\%2$, L^*H1 : $L^*H H\%1$ and L^*H2 : $L^*H H\%2$) and the corresponding “yardsticks” $Y_{.05}$ in each of the eight data sets obtained from 32 English listeners.

As is evident in Fig. 3, the contour-conditions largely fall into two groups, the fall-rise ($H^*L H\%1$ and $H^*L H\%2$) and the rise ($H^* H\%1$, $H^* H\%2$, $L^*H H\%1$, and $L^*H H\%2$). The estimated scale values of contours in the former group differed significantly from the estimated scale values of the contours in the latter group. Since lower estimated scale values correspond to higher preference scores, the two contour-conditions of the fall-rise were preferred to the contour-conditions of the rise. Within the rise contours, $H^* H\%1$ and $H^* H\%2$ had lower estimated scale values than $L^*H H\%1$ and $L^*H H\%2$, and hence were more favoured. The difference in preference between the $H^* H\%$ contours and the $L^*H H\%$ contours reached significance in data sets 1b, 2b, and 4b. As stimuli in these data sets represented the condition sentence-nonfinal continuation, it would seem that $H^* H\%$ was preferred to $L^*H H\%$ in particularly in the case of sentence-nonfinal continuation. Assuming that Cruttenden's

(1997: 50–52) rise nuclear tone (i.e. the initial movement from the nucleus is rise) is comparable to H* H% rather than L*H H%, of which the initial movement from the nucleus is low level, we may relate the preference for H* H% over L*H H% to the fact that H* H% is commonly used on sentence non-final IPs such as noun-phrase subjects and adverbials to signal non-finality in British English (Cruttenden 1997: 93–97).

Moreover, there was no significant difference in preference between the two phonetic variants of each contour. This indicates that the phonetic variants were judged to be equally appropriate, as expected.

Pearson's correlation coefficients (one-tailed) were obtained between the estimated scale values of the six contour-conditions in one data set and those in another data set. These correlations enable us to assess whether lexical content, pitch accent position and continuation type affected subjects' preference judgements. As is evident in Table 3, the estimated scale values obtained from one data set correlated significantly positively with those obtained from another data set. This indicates that the results of Scheffé's ANOVAs hold true across source expressions, continuation types and pitch accent positions. Hence, it can be concluded that in British English, H*L H% is the preferred continuation contour and between H* H% and L*H H%, H* H% is more favoured, independent of pitch accent position, continuation type and lexical content of the clause, though the preference for H* H% over L*H H% appears to be stronger in the signalling of sentence-nonfinal continuation.

Table 3. *Pearson’s correlation coefficients between estimated scale values in the eight data sets obtained from 32 British English listeners. The correlations are all significant at the 0.05 level.*

	1a	2a	3a	4a	1b	2b	3b	4b
1a	–							
2a	0.99	–						
3a	0.97	0.99	–					
4a	0.96	0.99	1	–				
1b	0.93	0.95	0.97	0.96	–			
2b	0.98	0.98	0.99	0.98	0.97	–		
3b	0.98	0.99	0.99	0.97	0.96	0.99	–	
4b	0.95	0.97	0.98	0.98	0.99	0.99	0.97	–

2.5.2. German data

Table 4 gives an overview of the estimated scale values per contour-condition, the corresponding “yardsticks” $Y_{.05}$, and the effect sizes of significant order effects in German data. The hypothesis of subtractivity was accepted in all data sets except for data set 4b. The order effects were significant in data sets 2a and 4b but had rather small effect sizes ($\eta^2 < 0.2$). The largest and the smallest estimated scale values (in bold) did not differ by more than the corresponding “yardstick” $Y_{.05}$ in data set 3b in German data. There were thus no significant overall differences in preference among the contour-conditions in this data set.

Table 4. *Estimated scale values, “yardsticks” $Y_{.05}$, and effect sizes of significant order effects in each set of German data. The content in each role is indicated in the leftmost column. The largest and the smallest estimated scales values in each data set are in bold.*

Continuation type	Final				Non-final			
	Final		Non-final		Final		Non-final	
Pitch accent position	1a	2a	3a	4a	1b	2b	3b	4b
Data set	1a	2a	3a	4a	1b	2b	3b	4b
$\hat{\alpha}_{H^*H\%1}$	-0.24	0.02	-0.01	-0.09	0.18	0.17	0.14	0.13
$\hat{\alpha}_{H^*H\%2}$	-0.27	0.1	0.21	-0.16	-0.32	0.24	-0.14	0.01
$\hat{\alpha}_{H^*LH\%1}$	-0.18	-0.29	-0.21	0.14	0.4	-0.15	0.04	-0.26
$\hat{\alpha}_{H^*LH\%2}$	0.26	-0.18	-0.29	-0.14	0.26	0.13	-0.05	0.19
$\hat{\alpha}_{L^*HH\%1}$	0.23	0.05	0.09	-0.02	-0.02	0.09	0.06	-0.01
$\hat{\alpha}_{L^*HH\%2}$	0.21	0.29	0.21	0.27	-0.49	-0.48	-0.05	-0.05
“yardstick” $Y_{.05}$	0.43	0.43	0.43	0.43	0.42	0.48	0.46	0.42
Order effects $\delta_{.05}$	n.s.	*	n.s.	n.s.	n.s.	n.s.	n.s.	*
		$\eta^2 = 0.17$						$\eta^2 = 0.11$

In the data sets where a significant overall difference in preference was found, the two items in many a contour-condition pair did not differ significantly in preference.¹⁰ It is therefore not illuminating to plot the estimated scale values as was done for the English data. Instead, the means of the estimated scale values were calculated for each of the six contour-conditions, following Rietveld and Gussenhoven (1987). We postpone further discussion on the preferred continuation contour in German to section 2.5.4 after the findings obtained from the Dutch data.

¹⁰ It is worth mentioning that one or both of the variants of H*L H% were rated considerably less favourably than one of the variants of L*H H% or H* H% in 6 out of the 8 cases where the difference in preference between contour conditions reached significance in data sets representing the IP-final pitch accent position (i.e. data sets 1a, 2a, 1b and 2b). This finding accords with Grabe’s (1998) observation that H*L H% does not occur on IP-final position in standard German.

2.5.3. Dutch data

Table 5 gives an overview of the estimated scale values per contour-condition, the corresponding “yardsticks” $Y_{.05}$, and the effect sizes of significant order effects in Dutch data. The hypothesis of subtractivity was accepted in all data sets except for data set 3b. The order effects were significant in all data sets except for data set 3a, but had small effect sizes ($\eta^2 < 0.3$). There were no significant overall differences in preference among the contour-conditions in data sets 2a and 4b as the largest and the smallest estimated scale values (in bold) did not differ by more than the corresponding “yardstick” $Y_{.05}$.

Similar to the German data, the two items in a large number of contour-condition pairs did not differ significantly in preference in the data sets with a significant overall difference in preference. Again, the means of the estimated scale values were calculated for each contour-condition. Further discussion on the preferred continuation contour in Dutch is available in the following section.

Table 5. *Estimated scale values, “yardsticks” $Y_{.0}$, and effect sizes of significant order effects in each set of Dutch data. The content in each role is indicated in the leftmost column. The largest and the smallest estimated scales values in each data set are in bold.*

Continuation type	Final				Non-final			
	Final		Non-final		Final		Non-final	
Pitch accent position	1a	2a	3a	4a	1b	2b	3b	4b
$\hat{\alpha}_{H^*H\%1}$	-0.02	0.16	0.15	-0.03	0.01	0.1	0.01	0.03
$\hat{\alpha}_{H^*H\%2}$	-0.07	-0.23	-0.2	-0.06	-0.33	-0.38	-0.18	0.09
$\hat{\alpha}_{H^*LH\%1}$	-0.29	0.08	-0.16	0.07	-0.03	0.25	0.06	0.05
$\hat{\alpha}_{H^*LH\%2}$	0.33	-0.06	-0.34	0.12	0.32	-0.28	-0.25	-0.15
$\hat{\alpha}_{L^*HH\%1}$	0.03	0.09	0.47	0.18	-0.04	0.18	0.36	-0.18
$\hat{\alpha}_{L^*HH\%2}$	0.02	-0.04	0.08	-0.28	0.07	0.14	-0.01	0.15
“yardstick” $Y_{.05}$	0.44	0.44	0.42	0.40	0.46	0.44	0.38	0.44
Order effects $\delta_{.05}$	*	*	n.s	*	*	*	*	*
	$\eta^2 = 0.16$	$\eta^2 = 0.14$		$\eta^2 = 0.2$	$\eta^2 = 0.12$	$\eta^2 = 0.14$	$\eta^2 = 0.28$	$\eta^2 = 0.21$

2.5.4. Preferred continuation contour in English, German and Dutch

Fig. 4 displays the mean estimated scale values for each contour-condition for all the three listener groups. Recall that lower scale values correspond to higher preference scores. In English listeners’ ratings, the pattern emerging from the mean estimated scale values accords with the results obtained from Scheffé’s ANOVAs, i.e. H*L H% was strongly preferred to H* H% and L*H H%; between H* H% and L*H H%, H*H% was more favoured; the two variants of each contour differed little in preference. This suggests that the mean scale values can serve as a reliable indicator of listeners’ preference among the six contour-conditions.

As can be seen in Fig. 4, Dutch listeners appeared to be most in favour for the high variant (i.e. H% is higher) of H* H% and, only marginally in favour for the high variant of H*L H%, but showed least preference for the low variant of L*H H%. The rating of H*L H% is not in accordance with 't Hart and Cohen's (1973) finding that H*L H% occurred frequently at the clause boundary. The discrepancy between their finding and our result may be due to the difference in the speech material. In 't Hart and Cohen (1973), subjects were asked to read proverbs and sentences selected from prose, while listeners in Experiment 1 rated semantically unsurprising sentences. It is likely that proverbs and prose tend to be read with intonation (e.g. H*L H% at a clause boundary) that is not necessarily frequently used for ordinary sentences like our stimuli.

Moreover, German listeners made only marginal differences among the six contour-conditions; H*L H% was relatively more favoured than the other contours. When plotting the mean estimated scale values for IP-final position and IP-nonfinal position separately, we noted that the preference for H*L H% was clearly observable on IP-nonfinal position but not so clear on IP-final position, as have been seen in Table 3 (see Footnote 11). These results seem to be in line with the observation (Féry 1993, Grabe 1998) that H*L H% is uncommon on IP-final position in Standard German, but argue against the earlier finding (Delattre 1965, Grover et al. 1987, Féry 1993) that German has a clear preference for rising contours at the clause boundary, in particular, L*H %. However, in Experiment 1, contours ending with % were not included. It may be speculated that a contour without a boundary tone, specifically the level high tone (H* %) or the half-completed rise (L*H %), is the preferred grammaticalised form for the continuation rise in German and its presence plays a decisive role in listeners'

preference judgements. Consequently, in the absence of contours ending with %, rising contours with H% are not rated more favourably than the fall-rise contour.

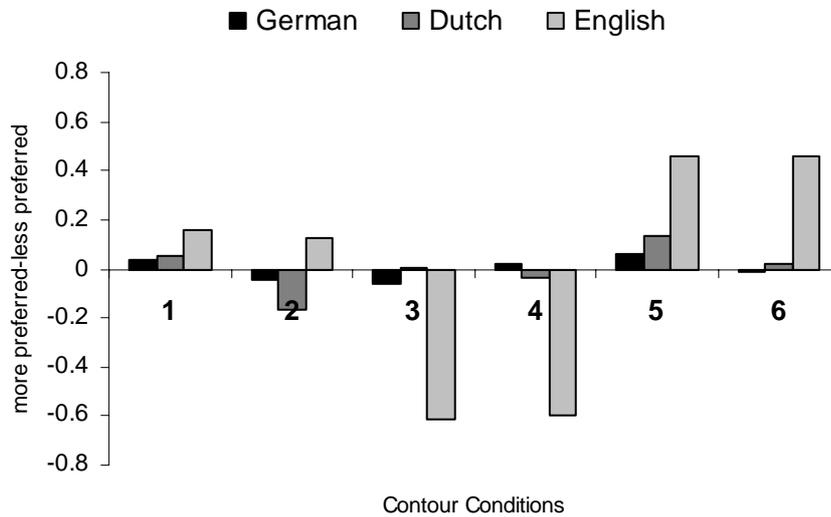


Fig. 4. Mean estimated scale values for each of the six contour-conditions (1. $H^*H\%1$, 2. $H^*H\%2$, 3. $H^*LH\%1$, 4. $H^*LH\%2$, 5. $L^*HH\%1$, and 6. $L^*HH\%2$) in British English, German and Dutch. A lower mean estimated scale value corresponds to a higher preference score.

To sum up, our results clearly point to a difference among British English, German and Dutch. Evidence from both Scheffé's ANOVAs and the mean estimated scale values convincingly shows that $H^*L H\%$ is the preferred continuation contour in British English, as we expected. Unexpectedly, there is also a difference in preference between $H^* H\%$ and $L^*H H\%$ with $H^* H\%$ being noticeably more favoured, in sentence-nonfinal continuation in particular. The comparison of the mean estimated scale values among the six contour-conditions suggests (a) that in German rise and fall-rise contours do not differ much and $H^*L H\%$ seems to entertain a slight preference;

and (b) that in Dutch H* H% is by and large more favoured than H*L H% and L*H H%; between H*L H% and L*H H%, H*L H% is more favoured.

3. Experiment 2: testing Hypothesis 2

The aim of experiment 2 was to establish how the difference in the preferred continuation contour(s) can affect the perception of final rise as a continuation cue. The hypothesis (Hypothesis 2) to be tested here was proposed on the basis of Hypothesis 1 that British English listeners would prefer H*L H%, German listeners would prefer L*H H%, and Dutch listeners would have no clear preference between L*H H%, H*L H% and H* H%. Under the same assumption that listeners are most sensitive to variations in the final rising portion of their preferred continuation contour, we adjusted Hypothesis 2 in the light of the findings of Experiment 1 as follows:

Adjusted Hypothesis 2:

- (1) When the final rise is part of a rise contour that is comparable to H* H%, British English and Dutch listeners will perceive a larger difference in continuation likelihood for a given interval of final rises than German listeners;
- (2) when the final rise is part of a fall-rise contour, British English listeners will perceive a larger difference in continuation likelihood than German and Dutch listeners.

3.1. Method

3.1.1. Stimuli

The hypothesis can be tested either by presenting listeners with stimuli in their native language, following Chen, Gussenhoven and Rietveld (2004), or by presenting listeners

with stimuli in an unknown language, following Gussenhoven and Chen (2000). In both cases, the final rising portion would be manipulated by, for example, varying the end pitch height in equal steps. As regards the option of using stimuli in listeners' native language, ideally we would want to use the same pitch interval between steps and include the same number of steps for every contour across languages in order to obtain comparable data. In practice this is very hard to do, because there is not sufficient information available to help to decide on the range of end pitch height. As a consequence, the chosen range of end pitch height may be acceptable for one contour in one language but not or less acceptable for another contour in the other languages or trigger connotations other than continuation. For this reason, we decided to use stimuli in an unknown language, where the chosen range of end pitch height would be considered a feature of this language and therefore by definition acceptable.

Eight utterances (see Appendix 2) were selected from the fourteen CVCVCV utterances with a sonorant for the second C (e.g., *mɛ wo nɐ*) designed by Gussenhoven and Chen (2000). The selected utterances were then read by the same speaker as in Experiment 1. The speaker was instructed to read each utterance with a rise contour on each syllable. The recording followed the same procedure as in Experiment 1. Best readings of the eight utterances were selected and digitised at a 32-kHz sampling rate. Speech manipulation was performed again by means of Praat.

Two sets of stimuli were generated. In one set of the stimuli, a rise contour comparable to H* H% realised on IP-final position was assigned to each syllable in each utterance. The rise contour started at the CV boundary and continued till the end of the vowel. For non-utterance final rises, the onset pitch value was 80 Hz and the offset pitch value was 100 Hz. For utterance-final rises, the onset pitch value was also 80 Hz;

the end pitch value was varied from 100 Hz to 180 Hz in 20-Hz steps, as illustrated in Fig. 5. This gave us 40 stimuli (8 source utterances \times 5 end pitch values), referred to as the Rise stimulus set.

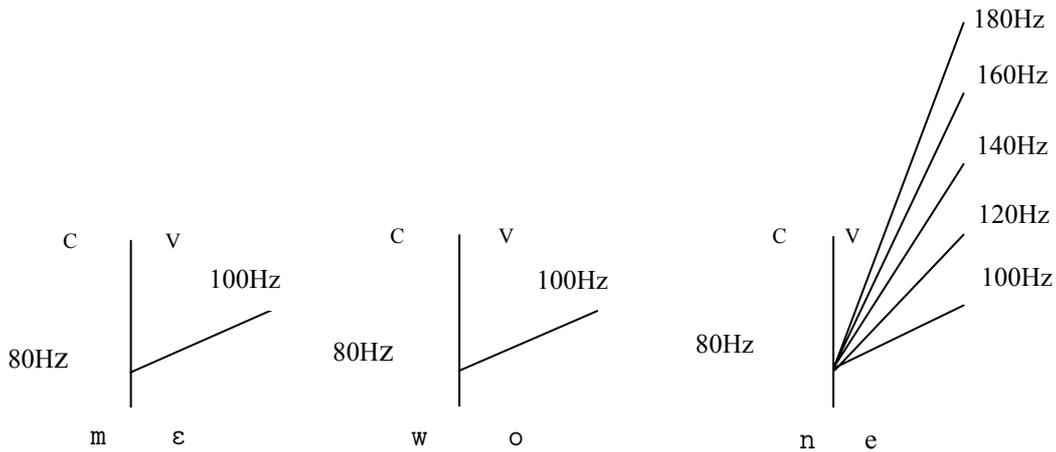


Fig. 5. Schematic representation of the sequence of rises imposed on the utterance ‘*mε wo ne*’. The end pitch of the utterance-final rise is varied in 5 steps.

The second set of the stimuli differed from the first set in that a fall-rise comparable to H*L H% realised on IP-final position was assigned to the utterance-final syllable. The fall started at the CV boundary and was 100-ms long; the rise continued till the end of the vowel and was varied from 100 Hz to 180 Hz in 20-Hz steps, as illustrated in Fig. 6. This gave us another 40 stimuli (8 source utterances \times 5 end pitch values), referred to as the Fall-Rise stimulus set.

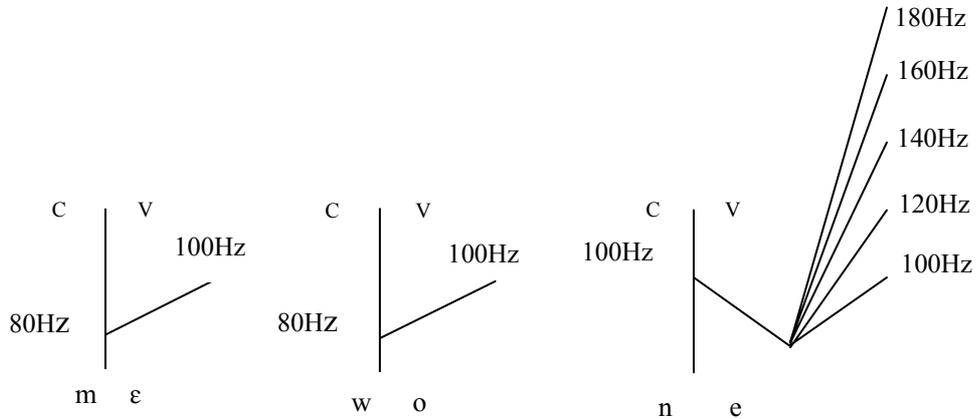


Fig. 6. Schematic representation of the sequence of rise_rise_fall-rise imposed on the utterance 'mε wo ne'. The end pitch of the utterance-final rise is varied in 5 steps.

These 80 stimuli were mixed manually and then divided into eight blocks of ten. To minimise order effects, two stimulus orders, order 1 and order 2, were produced by randomising the stimuli in each block and the blocks as a whole twice. Each stimulus order was recorded onto DAT tape (48 kHz in 16 bits) with a 4.5-s pause between stimuli, a 7-s pause between blocks, and a 200-ms 300-Hz sine wave preceding each block (to signal the beginning of a block), and was copied to a TDK audio tape. This gave us two eight-minute test tapes.

3.1.2. Procedure

Subjects who took part in Experiment 1 participated in Experiment 2 under equivalent circumstances. Subjects who received stimuli in the order AB in Experiment 1 were assigned test tape order 1 and those who received stimuli in the order BA to test tape order 2. The stimuli were presented to both groups of listeners as taken from Miao, a

little-known language spoken on an island off the Chinese coast.¹¹ They were told via written instructions in their native language that on each trial they would hear one utterance, which was selected from a sequence of two utterances on a single topic. Their task was to judge for each stimulus how likely it was said as the first utterance of the sequence of two. Subjects were asked to record their judgement for each stimulus by drawing a slash on a 100-mm horizontal line. The left end of the line was labelled as UNLIKELY and the right end MOST LIKELY. This scale is known as the Visual Analogue Scale (Wewers and Lowe 1990, Kreiman, Gerratt, Kempster, Erman and Berke 1993). It has been successfully used in previous research (e.g., Chen et al. 2004) on the perception of intonational meaning.

Prior to the experiment, four sequences of two utterances generated from expressions comparable to the source utterances for the stimuli were played to subjects as examples of Miao. This was to provide them with an idea of how continuation was signalled in Miao and to minimize any bias that was not related to their native language. Each syllable of the first utterance of each utterance sequence was assigned either a fall-rise or a rise and the end pitch value was varied from 120 Hz to 180 Hz. The second utterance of each utterance sequence ended with a fall (falling from 100 Hz to 80 Hz) or a weak rise (rising from 80 Hz to 90 Hz).

Following the examples, subjects were given a practice session, in which they listened to four utterances generated from another four nonsense utterances comparable to the source utterances for the stimuli and performed the experimental task.

¹¹ Note that the Miao in our experiment is not the same as the Miao that is spoken by the Mao ethnic group from Yunnan province in China. We borrowed the name of an existing language to refer to our made-up utterances because we wanted to present the utterances as taken from a language spoken in China.

3.2. Statistical analysis and results

In order to make inferences about the findings of Experiment 2 from those of Experiment 1, only data obtained from the subjects whose data were included into the analyses of Experiment 1 were taken into account for the analysis here. A repeated measures ANOVA was performed on the data at the significance level of 0.05 with the ‘likelihood’ score as the dependent variable. The analysis included three within-subject factors, Contour Shape of the utterance final syllable (2 levels: rise, fall-rise), End Pitch of the utterance-final rise (5 levels), utterance (8 levels), and one between-subject factor Language Background (3 levels). Where sphericity was violated, the Huynh-Feldt corrected p-values are used.

The main effects for the variables End Pitch and Contour Shape were found to be significant. The main effect of End Pitch ($F_{4, 208} = 100.5$, $p < 0.001$, $\eta^2_{\text{partial}} = 0.7$) is such that an utterance with a higher end pitch was judged as more likely to be said as the first utterance of a sequence of two. The main effect of Contour Shape ($F_{1, 60} = 88.21$, $p < 0.001$, $\eta^2_{\text{partial}} = 0.6$) shows that the rise contour was associated with a higher likelihood of continuation than the fall-rise contour. The main effect of Utterance was also found to be significant ($F_{7, 364} = 2.213$) but with a very small effect size ($\eta^2_{\text{partial}} = 0.04$). It appeared that stimuli derived from source utterances 4 and 7 sounded marginally less likely to be continued than stimuli derived from the other source utterances.

Significant interactions were found between the within-subject factors and the between-subject factor Language Background as well as between the within-subject factors. The effect of the three-way interaction of End Pitch \times Contour Shape \times Language Background ($F_{8, 208} = 2.23$, $p < 0.05$, $\eta^2_{\text{partial}} = 0.1$). is relevant to us and

considered in detail here. Fig. 7 displays the effect of the interaction between End Pitch and Language per contour shape. As can be seen in Fig. 7a, when the utterance-final syllable was said with a rise, British English and Dutch listeners were similar in their ratings; they differed from German listeners in that they perceived a substantially larger meaning difference for the given interval of end pitch heights. When the utterance-final syllable was said with a fall-rise, German and Dutch listeners were similar in their ratings; they differed from British English listeners in that they perceived a smaller meaning difference for the given interval of end pitch heights, as shown in Fig. 7b.

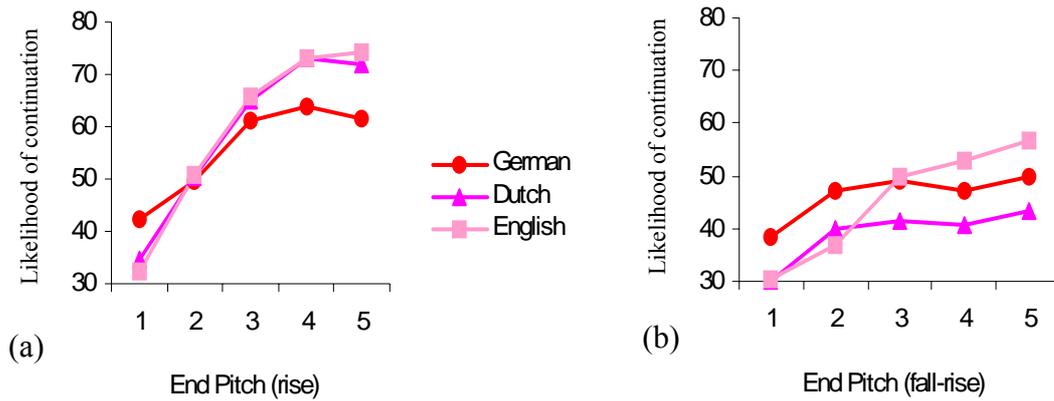


Fig. 7. Mean scores for likelihood of continuation of British English listeners ($n = 32$), German listeners ($n = 16$) and Dutch listeners ($n = 16$) in the 5 end pitch height conditions of the utterance-final rise (panel a) and the utterance-final fall-rise (panel b). End Pitch Height increases from level 1 to level 5.

3.3. Discussion

As expected, when the utterance-final syllable had a fall-rise contour, British English listeners, judging H*L H% as their preferred continuation contour, perceived a larger difference in continuation likelihood for the given interval of end pitch heights than

German and Dutch listeners, who showed no preference for H*L H%. When the utterance-final syllable was said with a rise contour, Dutch listeners, who rated H* H% more favourably than H*L H% and L*H H%, and British English, who rated H* H% noticeably more favourably than L*H H%, perceived a larger meaning difference than German listeners, who rated the three contours similarly, again as predicted. Unexpectedly, all groups of listeners perceived a larger meaning difference in utterance-final rise than in utterance-final fall-rise. This indicates a strong language-independent bias towards the final rise, which showed up even when the listeners were familiarised with both final rise and final fall-rise as cues to continuation.

4. Summary

In this study, we have examined first how British English, German and Dutch listeners differ in their preferred continuation pitch contour at the clause boundary (Experiment 1), and second how such differences in these languages affected the perception of final rises as cues for continuation in Miao.

With respect to the preferred continuation pitch contour, it was found that British English listeners have a clear preference for H*L H% over H* H% and L*H H%, independently of continuation type and pitch accent position, as expected. Between H* H% and L*H H%, the former is more favoured than the latter, in particular, as regards sentence-nonfinal continuation. This is argued to be related to the common use of H* H% on sentence non-final IPs such as noun-phrase subjects and adverbials to signal non-finality in British English (Cruttenden 1997). German listeners distinguish only marginal differences among the three contours and seem to have a slight preference for H*L H%. This finding should not be interpreted to mean that all contours with a

continuation rise are equally favourable at the clause boundary to signal continuation in German. It was suggested that either the high level tone (H* %) or the half-completed rise (L*H %) may be the preferred grammaticalised form for continuation rise in German and decide listeners' preference judgments for a contour. In the absence of contour ending with %, rising contours with H% and the fall-rise contour H*L H% are thus rated similarly. By and large, Dutch listeners prefer H* H% to the H*L H% and L*H H%. The discrepancy between our results and 't Hart and Cohen's (1973) finding that H*L H% occurs frequently at the clause boundary was attributed to the difference in speech material: subjects in the former rated semantically unsurprising sentences while subjects in the latter read proverbs and sentences selected from prose.

In the perception of final rises as continuation cues in a made-up language Miao, we found that although all the three groups of listeners largely associate a higher end pitch with a higher degree of likelihood of continuation, the perceived meaning difference for the given interval of end pitch heights varies with the contour shape of the utterance final syllable. When the utterance ends with a fall-rise contour, British English listeners perceive a larger difference in continuation likelihood for the given interval of end pitch heights than German and Dutch listeners. This is consistent with the finding that H*L H% is the preferred continuation contour in British English. When the utterance ends with a rise contour, British English and Dutch listeners distinguish a sharper meaning difference for the given interval of end pitch heights than German listeners. Again, this difference can be related to how favourably H* H% is rated in British English and Dutch.

These findings suggest a couple of topics for further investigation. First, because rise contours with a half-completed rise (i.e. H* % and L*H %) have also been reported

to be used to signal continuation in German and Dutch, we do not know whether the preference identified in Experiment 1 for these two languages would still hold if H* % and L*H % were included among the candidates. There is already an indication that German may favour contours ending with % over contours ending with H%. A similar experiment in which all five contours (H* H%, H*L H%, L*H H%, H* % and L*H %) are included would be desirable to further establish the preferred continuation contour(s) in German and Dutch. Second, it has been shown (e.g., Blaauw 1995) that read speech and spontaneous speech differ in the signalling of continuation at clause boundaries. As the utterances in Experiment 1 may sound more like read speech than spontaneous speech, the present findings on preferred continuation pitch contour may not necessarily hold for spontaneous speech. It would be useful to verify our findings on stimuli that are representative of spontaneous speech.

Acknowledgements: I would like to thank Toni Rietveld for useful comments on the experimental design, Gerard Latjes and Christiane Ulbrich for administering the Dutch part and the German part of the experiment, Bob Ladd and colleagues at the University of Edinburgh for helping recruit English subjects, Jean-Pierre Sprengers for helping measure the data, Ellen Bard, Mika Ito and Toni Rietveld for statistical advice, Frauke Hellwig and Marieke Hoetjes for making the figures and tables, and Jörg Peters and Toni Rietveld for helpful discussion and comments on earlier versions of this text. Finally, I am grateful to Bob Ladd, Marc Swerts, Vincent van Heuven, and an anonymous reviewer, whose detailed comments led to considerable improvements in this paper.

References

Arvaniti, Amalian and Mary Baltazani 2004 Intonational analysis and prosodic annotation of Greek spoken corpora. In Sun-Ah Jun (Ed.), *Prosodic Models and Transcription: Towards Prosodic Typology*. Oxford: Oxford University Press.

Avgustinova, Tania., and Bistra Andreeva 1999 Thematic intonation patterns in Bulgarian clitic replication. In J. J. Ohala, Y. Hasegawa, M. Ohala, D. Granville, and A. C. Bailey (Eds.), *Proceedings of the 14th international congress of phonetic sciences*, 1501–1504. Berkely: University of California.

Beckman, Mary E., Manuel Díaz-Campos, Julia Tevis McGory, and Terrell A. Morgan 2002 Intonation across Spanish, in the Tones and Break Indices framework. *Probus* 14, 9–36.

Boersma, Paul, and David D. Weenink 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>.

Calhoun, Sasha 2003 The nature of Theme and Rheme Accents. Retrieved from <http://www.cstr.ed.ac.uk/~s0199920/research.html>.

Chen, Aoju, Carlos Gussenhoven, and Toni Rietveld 2004 Language-specificity in the perception of paralinguistic intonational meaning. *Language and Speech* 47 (4), 311–349.

Chorianopoulou, Elisabeth 2002 Evaluating prosody prediction in synthesis with respect to Modern Greek prenuclear accents, Master thesis, Linguistics, University of Edinburgh.

Cruttenden, Alan 1997 *Intonation* (2nd ed.). Cambridge: Cambridge University Press.

Delattre, Pierre 1963 Comparing the prosodic features of English, German, Spanish and French. *International Review of Applied Linguistics* 1, 193–210.

Delattre, Pierre 1965 *Comparing the phonetic Features of English, French, German and Spanish: An Interim Report*, 23–43. Heidelberg: Julius Groos Verlag.

Féry, Caroline 1993 *German Intonational patterns*. Tübingen: Niemeyer.

Grabe, Esther 1998 *Comparative intonational phonology: English and German*. PhD dissertation, University of Nijmegen. Published in Max Planck Institute Series in Psycholinguistics.

Grover, Cynthia, Donald G. Jamieson, and Michael. B. Dobrovolsky 1987
Intonation in English, French and German: perception and production. *Language and Speech* 30, 277–296.

Gussenhoven, Carlos 1984 *On the grammar and semantics of sentence accents*.
Dordrecht: Fortis publications.

Gussenhoven, Carlos 2002 Intonation and interpretation: phonetics and phonology. In B. Bel, and I. Marlien (Eds.), *Proceedings of the Speech Prosody*, 47–57. Aix-en-Provence: Université de Provence.

Gussenhoven, Carlos 2004 Transcription of Dutch Intonation. In Sun-Ah Jun (ed.), *Prosodic typology and transcription: A unified approach*. Oxford: Oxford University Press.

Gussenhoven, Carlos, Toni Rietveld, and Jacques Terken 1999 Transcription of Dutch Intonation: Courseware, version 1.1. URL: <http://lands.let.kun.nl/todi/todi/home.htm>

Gussenhoven, Carlos, and Aoju Chen 2000 Universal and language-specific effects in the perception of question intonation. In B. Yuan, T. Huang, and X. Tang (Eds.), *Proceedings of the sixth international conference on spoken language processing*, Vol. I, 91–94. Beijing: China Military Friendship Publish.

Haan, Judith 2001 *Speaking of questions: an exploration of Dutch question intonation*. PhD dissertation. Utrecht: LOT.

Krieman, J., Gerratt, B. R., Kempster, G. B., Erman, A., and 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.

Ladd, Robert, D. 1996 *Intonational Phonology*. Cambridge: Cambridge University Press.

Lahiri, Aditi, and Jennifer Fitzpatrick-Cole 1999 Emphatic clitics and focus intonation in Bengali. In René Kager, and Wim Zonneveld (Eds.) *Phrasal Phonology*, 119–144. Nijmegen: University of Nijmegen Press.

Pierrehumbert, Janet B. 1981 Synthesizing intonation. *Journal of the Acoustical Society of America* 70 (4), 985–995.

Pierrehumbert, Janet B. and Mary E. Beckman 1988 *Japanese tone structure*. Cambridge, MA: MIT press.

- Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech, and Jan Svartvik 1980 *A grammar of contemporary English*. Essex: Longman Group Ltd.
- Rietveld, Toni and Carlos Gussenhoven 1985 On the relation between pitch excursion size and pitch prominence. *Journal of Phonetics* 13, 299–308.
- Rietveld, Toni and Carlos Gussenhoven 1987 Perceived speech rate and intonation. *Journal of Phonetics* 13, 273–285.
- Rietveld, Toni and Carlos Gussenhoven 1995 Aligning pitch targets in speech synthesis: Effects of syllable structure. *Journal of Phonetics* 23, 375–385.
- Sanderman, Angeliën A. and René Collier 1995 Prosodic phrasing at the sentence level. In F. Bell-Berti and L. J. Raphael (Eds.), *Producing speech: contemporary issues – for Katherine Safford Harris*, 321–332. AIP Press, Woodbury NY.
- Sanders, Margreet J. 1996 *Intonation contour choice in English*. PhD dissertation, University of Utrecht. Published in OTS Dissertation Series.
- Scheffé, Henry 1952 An analysis of variance for paired comparisons. *Journal of the American Statistical Association* 47, 381–400.
- 't Hart, Johan, and Antonie Cohen 1973 Intonation by rule: a perceptual quest. *Journal of Phonetics* 1, 309–327.

't Hart, Johan, Rene Collier, and Antonie Cohen 1990 *A perceptual study of intonation*. Cambridge: Cambridge University Press.

Vanvik, Arne 1979 Three functions of Intonation. Retrieved 1 September, 2004 from <http://folk.uio.no/janneto/300kopi/node14.html>.

Von Essen, Otto 1956 *Grundzüge der hochdeutschen Satzintonation*. Ratingen: Henn.

Wewers M. E. and Lowe N.K. 1990 A critical review of visual analogue scales in the measurement of clinical phenomena. *Research in Nursing and Health* 13, 227–236.

To appear in *Tones and Tunes, Volume II, Phonetic and Behavioural Studies in Word and Sentence Prosody*. Carlos Gussenhoven & Tomas Riad (eds.) Mouton de Gruyter, Berlin in: *Phonology and Phonetics*. Series editor: Aditi Lahiri

Appendix 1

Source expressions for the experimental stimuli in Experiment 1. Syllables bearing a nuclear pitch accent are in capitals, of which bearing the target nuclear pitch accents (i.e. H*L, H*, L*H) are underscored; syllables bearing pre-nuclear pitch accent are in italics.

English

- 1a. The *game* will take a YEAR. Nobody is INterested.
- 1b. The *game* will take a YEAR and *brings* in a lot of MOney.

- 2a. The *story* is too LONG. The *plot* is BORing.
- 2b. The *story* is too LONG but is *fun* to READ.

- 3a. I *ran* into WIlliam. We *had* a CHAT.
- 3b. I *ran* into WIlliam and *passed* him the MESSage.

- 4a. We *were* in RoMAnia. We *had* a great TIME.
- 4b. We *were* in RoMAnia but have *kept* our aPARTment here.

German

- 1a. Das *Spiel* dauert ein JAHR. Niemand intereSSIERT sich dafür.
- 1b. Das *Spiel* dauert ein JAHR und *bringt* viel *Geld* ein.

- 2a. Die *Geschichte* ist zu LANG. Der *Inhalt* ist LANGweilig.
- 2b. Die *Geschichte* ist zu LANG, aber *angenehm* zu LEsen.

- 3a. Ich *traf* WIlhelm. Wir *plauderten* ein WENig.
- 3b. Ich *traf* WIlhelm und *übergab* ihm die NACHricht.

- 4a. Wir *waren* in RuMÄnien. Das war unglaublich SCHÖN.
- 4b. Wir *waren* in RuMÄnien, haben jedoch unsere *Wohnung* hier beHAlten.

Dutch

- 1a. Het *spel* duurt een JAAR. Niemand is geïntereSSEERD.
- 1b. Het *spel* duurt een JAAR en *bren*gt veel GELD op.

- 2a. Het *verhaal* is te LANG. De *plot* is SAAI.
- 2b. Het *verhaal* is te LANG maar is *leuk* om te LEzen.

- 3a. Ik *zag* WIlliam. We *maak*ten een PRAAtje.
- 3b. Ik *zag* WIlliam en *gaf* hem de BOODschap door.

- 4a. We *waren* in Romenie. Het was geWELdig.
- 4b. We *waren* in Romenie maar hebben onze *flat* hier AANgehouden.

Appendix 2

Source utterances for the experimental stimuli in Experiment 2.

1. tɔ nɛ wɔ
2. dɔ mɑ lɛ
3. mɛ wɔ nɛ
4. lɑ jɔ rɛ
5. zɔ rɛ jɔ
6. tɛ mɛ wɔ
7. dɔ jɑ lɔ
8. zɛ nɔ wɛ