UNIVERSAL AND LANGUAGE-SPECIFIC EFFECTS IN THE PERCEPTION OF QUESTION INTONATION

Carlos Gussenhoven and Aoju Chen

Centre for Language Studies, University of Nijmegen,
Post box 9103, 6500 HD Nijmegen, The Netherlands

ABSTRACT

Three groups of monolingual listeners, with Standard Chinese, Dutch and Hungarian as their native language, judged pairs of trisyllabic stimuli which differed only in their pitch pattern. The segmental structure of the stimuli was made up by the experimenters and presented to subjects as being taken from a little-known language spoken on a South Pacific island. Pitch patterns consisted of a single rise-fall located on or near the second syllable. By and large, listeners selected the stimulus with the higher peak, the later peak, and the higher end rise as the one that signalled a question, regardless of language group. This result is argued to reflect innate, non-linguistic knowledge of the meaning of pitch variation, notably Ohala's Frequency Code. A significant difference between groups is explained as due to the influence of the mother tongue.

1. INTRODUCTION

Intonational melody is widely used for signalling discoursal meanings. Ladd identified two points of view: the *Strong Universalist Hypothesis,* according to which pitch rises signal questions and pitch falls signal statements, and the *Nuclear Tone Hypothesis,* according to which the distribution of contour types over functions is language-specific, determined by the grammar, and arbitrary [1]. The form-function correlation of the Strong Universalist Hypothesis is documented in [2,3], and supported in [4], which explained the connection between questions and pitch rises by the relatively high muscular tension in the larynx responsible for the creation of high pitch. In comparison with low pitch as found at the end of falls, high pitch was believed to be associated with higher degrees of excitement and suspense. An alternative explanation, based on the use of pitch variation in the animal kingdom, was offered in [5,6,7]. Other relevant treatments are [8,9,10].

Liberman first used the term 'intonational lexicon' for the notion of a set of pitch accents or nuclear tunes which have linguistic meaning, referred to as 'ideophonic' meaning [11]. In spite of the suggestion of iconicity carried by this term, the idea is that the lexicon is a set of grammatical forms, with language-specific meanings. Liberman's position is thus an exemplar of the Nuclear Tone Hypothesis. Support for it is found in languages that use rises for statements, like Belfast English [12] and Chickasaw [13], or falls for questions, like Roermond Dutch [14] and, again, Chickasaw, which patterns contradict the form-function relation on which the Strong Universalist Hypothesis is built. Ladd therefore concludes that of the two hypotheses, the Nuclear Tone Hypothesis is the more explanatory: while it can explain all form-function relations on the grounds that any such relation can be defined in the grammar, the Strong Universalist Hypothesis can only explain those that conform to the pattern of rising questions and falling statements.

A drawback of this conclusion is that it is no longer clear why the pattern of rising questions and falling statements is so widespread. One of the central tenets in linguistics is that form-function relations are arbitrary, and so a situation whereby the same form-meaning relations appear in a majority of largely unrelated languages is a cause for concern.

Recently, the first author proposed that the universal and the language-specific communicative effects derive from two different language components [15]. One component is the intonational lexicon, the set of morphemes invested with intonational meaning. For Chickasaw, for example, these include H* L% for 'interrogative' and H* H% for 'statement', and for Dutch these include H*L H% for 'interrogative' and H*L L% for 'statement'. The meaning and the phonological form of these morphemes will have to be learned by the language learner, and the relation between them is in principle arbitrary. (We here leave out of consideration the apparent fact that intonational meaning is restricted to discoursal meaning.)

The other component is the phonetic implementation module, whose nature differs considerably from that of the phonological component. Implementation concerns the phonetic interpretation of the discretely configured phonological surface representation. In itself, phonetic implementation is not universal: part of learning a language is learning how to pronounce your phonology [e.g.16]. Moreover, it has been claimed that phonetic implementation is under the control of the speaker, who exercises this control not only to his social advantage in the case of prestige-related variation, but also to enhance the phonological contrasts of his language [17]. However, this phonetic implementation module is also the place where speakers avail themselves of the opportunity to signal universal (non-grammatical) meanings of pitch variation.

How do speakers know what the universal (non-grammatical) meanings are and how they are to be signalled? These meanings are shared by our larger biological phylum, derive from biologically determined codes, and are thus innate, though non-linguistic. The most important of these is the Frequency Code [5,6,7], which is based on the fact that smaller larynxes produce higher notes than larger ones. One of its manifestations is that high pitch sounds vulnerable and submissive, while low pitch sounds protective and dominant. This is why high, and particularly high-ending utterances seem to sound dependent, appealing, questioning, etc., while conversely, low and low-ending utterances seem to sound authoritative, powerful, and assertive. A second natural form-function relationship, the Effort Code, is based on the fact that greater effort creates more elaborate, more explicit realisations [15]. In the context of pitch variation, greater effort corresponds with wider excursions, and as a result, a higher peak will sound more emphatic than a lower peak. Such meanings exist in language communication not because they are part of language, but because they belong to our non-linguistic biological inheritance. Morphemes with meanings that conform to these universal codes may have to be seen as grammaticalisations of the non-linguistic codes, but languages may change over time in ways that contradict them. Thus, the meanings of phonological forms *may conflict with* the universal meanings, while biases in the phonetic implementation *will always directly express* the universal meanings.

Both the Frequency Code and the Effort Code may give rise to secondary effects, due to a mechanical connection between peak height and peak alignment. In [15], it is speculated that, since rise-falls with higher peaks take a longer time to complete than rise-falls with lower peaks, later peaks may be used as substitutes for higher peaks. Late peaks may therefore signal emphasis (by the Effort Code), and - more importantly in the context of this paper - Interrogativity (by the Frequency Code).

Thus, we predict that all humans, regardless of language background, can capitalise on their innate non-linguistic knowledge when judging whether an utterance is intended as a question. If listeners without knowledge of the intonation system of the language concerned are presented with two utterances, each with a rising-falling contour, and are asked to pick the one that most probably is a question, they will select the one with the higher peak, or with the later peak, or with higher-ending pitch. The purpose of this paper is to demonstrate that this is so. Of course, people's judgements will naturally be influenced by experience, and responses may therefore be biased towards the patterns that are used for question intonation in the languages they are familiar with.

2. A CROSS-LINGUISTIC PERCEPTION EXPERIMENT

In order to show that humans, regardless of language background, can avail themselves of the innate knowledge embodied in the Frequency Code, we selected three languages in which Interrogativity is expressed differently, *viz.* Dutch, Hungarian and Standard Chinese (Mandarin). In addition to a syntactic interrogative construction, Dutch has a number of question intonation contours that end with a final rise (H%). Hungarian has no interrogative syntactic construction, and expresses interrogativity by means of a L* pitch accent followed by a HL%-boundary sequence, which creates a pitch peak on the final syllable if the accenting L* occurs on the penultimate syllable, which has low pitch. This pattern contrasts with H* L%, used for declarative intonation, which is realised by a pitch peak on the accented syllable, followed by low pitch. Thus, if the accent occurs on the penultimate syllable, the pitch peak is early in statements and late in questions [18,19]. Higher peaks were earlier found to be a cue to interrogativity for Hungarian listeners, in addition to later peaks [19]. Chinese employs a number of sentence-final question particles, has no specific question intonation, but may raise pitch for questions [20,21].

We also selected three variables which could be related to the Frequency Code: peak height, peak alignment (the secondary effect identified above), and final pitch.

2.1 Stimuli

Fourteen CVCVCV-expressions with a sonorant for the second C were made up which were to serve as source utterances for our stimuli. We avoided high vowels to reduce intrinsic F0 effects. Three phonetically trained female speakers of Dutch were recruited to record the source utterances on digital audio tape (DAT) in a sound-treated room. They were given phonetic transcriptions, and were asked to read the utterances as smoothly as possible with a non-emphatic stress on the penultimate syllable. We selected the best readings by the best speaker, and digitised these utterances at 32 kHz. Manipulation was performed with the help of Praat [http://fonsg3.let.uva.nl/praat/]. First, we normalised the duration of the stressed (penultimate) vowel in all source utterances to 120 ms. The peak, which was realised as a 30 ms high plateau preceded by a 120 ms rise and followed by a 120 ms fall, was subsequently varied from

250 Hz to 370 Hz in 30 Hz-steps in each source utterance (H1 to H5). Each of the peak height values was then combined with five peak alignments by shifting the rise-plateau-fall contour through the vowel of the accented syllable in 30ms-steps, starting from a position with the onset of the high plateau at the CV boundary (A1 to A5). The final F0 was 140 Hz; other values were fixed as in Figure 1. These 25 combinations of peak alignment and peak height were distributed over five source utterance pairs in a Latin square to create 50 stimuli. Each stimulus had a different source utterance from its counterpart with the same F0 contour, but had the same vowel in the stressed syllable. This stimulus set will be referred to as the L%-set.

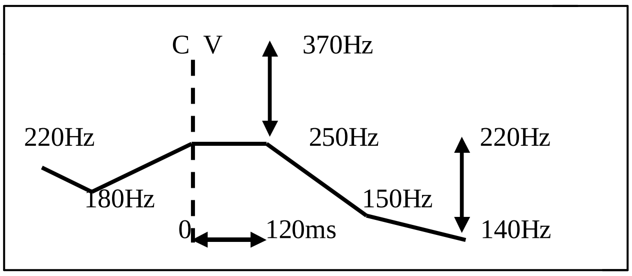

Figure 1. An artificial pitch contour with indication of the ranges of peak height, peak alignment, and end pitch.

A second (partly overlapping) set of ten source utterances was selected, such that five pairs with the same final vowel could be formed. Each pair was provided with one of five values for the final pitch, from 140 Hz to 220 Hz in steps of 20 Hz. These ten versions were combined with the three combinations of peak alignment and peak height A1H1, A3H3, A5H5, so that we obtained 30 stimuli, a set referred to as the H% set.

Each of the 80 stimuli was combined with a segmentally identical anchor stimulus which had a peak with average peak alignment and peak height (A3H3) and a 140Hz final pitch, so as to form 80 experimental stimulus pairs. This was done to enable listeners to choose the question from two intonational versions of each expression. In order to minimise the risk that listeners would detect the fact that there was an anchor, we added 80 filler pairs. The manipulations of filler stimuli differed from those of experimental stimuli in the range of peak height, the range of peak alignment, utterance onset F0 value and final pitch. Experimental stimulus pairs and filler pairs were randomised and divided into 10 blocks of 16 pairs, and presented to the three groups of listeners in equivalent circumstances. Each group consisted of 30 linguistically naive subjects between 18 and 30 years old, who were instructed to pay attention to the intonation of the stimuli, and asked to judge which of the two stimuli in each pair sounded more like a question.

3. RESULTS

Two analyses of variance were performed on the data in the L%-set, with the factors Peak Height, Peak Alignment and Language, as well as those in the H%-set, with the factors End Pitch, Peak Condition, and Language. We report p-values after correction for sphericity (Huynh-Feldt ε). In the L%-set, there were significant interactions between Peak Height and Peak Alignment ($F_{16,1392}=3.14$, $p<0.01$) and between Peak Height and Language ($F_{8,348}=8.90$, $p<0.01$). The first of these is of no interest in our context: it appears that the effect of each of these variables is stronger when combining with lower values for the other. The second interaction is due to the lesser sensitivity of the Chinese listeners to the height of the peak. This is shown in Figure 2.

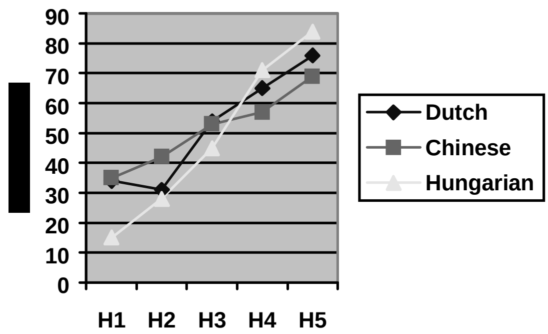

Figure 2. Percentage "Question" judgements as a function of peak height by three groups of listeners with ordinal interaction between listeners' language and peak height.

Main effects were found for Peak Height ($F_{4,348}=148.40$, $p<0.01$) and Peak Alignment ($F_{4,348}=12.51$, $p<0.01$). The effect of Peak Alignment was such as to make later peaks sound more like questions; it was weaker than that for Peak Height: percentages from A1 to A5 are 43, 49, 52, 53 and 57.

In the H%-set, we found a significant interaction between Language and Peak Condition ($F_{4,174}=13{,}86$, $p<0.01$), and significant main effects for End Pitch ($F_{4,348}=30.58$, $p<0.01$), and for Peak Condition ($F_{2,174}=167{,}47$, $p<0.01$). The interaction between Language and Peak Condition is probably the same effect as that found between Language and Peak Height in the L%-set: the Hungarian listeners seem more impressed by the peak condition than the Dutch listeners and in particular the Chinese listeners, as shown in Figure 3. The effect of End Pitch is such that the higher it is, the more likely judges are to opt for questions, though it was smaller than we expected (respectively, 49, 49, 65, 67 and 68 per cent).

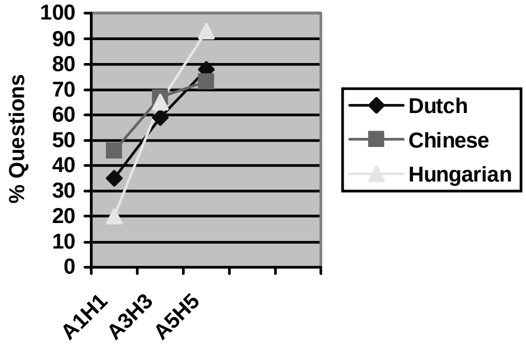

Figure 3. Percentage "Question" judgements as a function of peak condition by three groups of listeners, with ordinal interaction between Language and Peak Condition.

4. DISCUSSION

Higher peaks, later peaks, and higher end pitch lead Dutch, Chinese and Hungarian listeners to believe that utterances in an unknown language are more likely to be questions. Prosodically, these three languages express interrogativity differently, and since they scored stimuli in a language they could not have been familiar with (it was made up by the experimenters), these results suggest that these cues are universal. It has been argued that this universal knowledge is non-linguistic, and was identified by Ohala [5,6] as the Frequency Code. The fact that Hungarian speakers appear to be more strongly influenced by peak height than Chinese listeners shows that language background is an active component in listeners' interpretation strategies, in addition to the universal, non-linguistic strategy.

Acknowledgement. We would like to thank István Bátfai, Gerard Latjes and Yunpeng Zu for their help with the administration of the tests, and Toni Rietveld for allowing us to freely call on his statistical expertise.

5. REFERENCES

1. Ladd, D.R. "On Intonational Universals". In Myers, T. et al. (eds). *The Cognitive Representation of Speech*, North Holland Publishing, Amsterdam, 389-397, 1981.

2. Hermann, E. *Probleme der Frage,* VandenHoeck & Ruprecht, Göttingen, 2 vols., 1942.

3. Bolinger, D. "Intonation across languages". In Greenberg, J. (ed.) *Universals of Human Language. Vol II: Phonology,* Palo Alto, CA: Stanford University Press. 471-524, 1978.

4. Liberman, P. *Intonation, Perception, and Language,* MIT Press, Cambridge MA, 1967.

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

6. Ohala, J.J. "An ethological perspective on common cross-language utilization of Fo in voice," *Phonetica* 41: 1-16, 1984.

7. Ohala, J.J. "The frequency code underlies the sound-symbolic use of voice pitch". In Hinton, L. et al. (eds.) *Sound symbolism*, Cambridge University Press, 1994

8. Cruttenden, A. "Falls and rises: meanings and universals," *Journal of Linguistics* 17: 77-91, 1981.

9. Lindsey, G. *Intonation and interrogation: Tonal structure and the expression of a pragmatic function in English and other languages*, UCLA, 1985.

10. Bolinger, D. *Intonation and its Parts,* Stanford University Press, Stanford CA, pp. 194 ff., 1986.

11. Liberman, M. Y. *The Intonational System of English*, Garland publishing, New York.

12. Jarman, E.& Cruttenden, A. "Belfast intonation and the myth of the fall," *JIPA* 6: 4-12, 1976.

13. Gordon, M.K. "The intonational structure of Chickasaw." *ICPhS* 14, 1993-1996, 1999.

14. Gussenhoven, C. "The boundary tones are coming: On the non-peripheral realization of intonational boundary tones". In M.B. Broe & J.B. Pierrehumbert, *Papers in Laboratory Phonology V.* Cambridge University Press, 132-151, 2000.

15. Gussenhoven, C. "Why questions intonations rise and why they sometimes fall". In Hoey, M & Scott, N. (eds) *'Questions' Conference*. University of Liverpool Press, forthcoming.

16. Pierrehumbert, J.B. "Phonological and phonetic representation," *Journal of Phonetics* 18: 375-394.

17. Kingston, J. & Diehl, R.L. "Phonetic knowledge," *Language* 70: 419-454.

18. Ladd, D.R. *Intonational Phonology*. Cambridge University Press, Cambridge, 1996.

19. Gósy, M. & Terken, J. "Question marking in Hungarian: Timing and height of pitch peaks". *Journal of Phonetics* 22**:** 269-281, 1994.

20. Kratochvil, P. "Intonation in Beijing Chinese," in Hirst, D. & Di Cristo, A. *Intonation Systems: A Survey of Twenty Languages*. Cambridge University Press, 417-431.

21. Shen, X.S. *The Prosody of Mandarin Chinese*. UC Berkeley, 1990.