Speech Prosody 2006
Dresden, Germany
May 2-5, 2006

# Phrase-Final Pitch Discrimination in English

*Fred Cummins*[1] *& Colin Doherty*[2] *& Laura Dilley*[3]

[1]University College Dublin
[2]Royal College of Surgeons in Ireland
[3]The Ohio State University

fred.cummins@ucd.ie, cpdoherty@partners.org, dilley.28@osu.edu

## Abstract

We investigate the discrimination of phrase final pitch contours within a continuum from statement to question in English. Previous work in German [14] and Dutch [13] has raised questions about the relationship between discrimination sensitivity and category structure within this continuum. To clarify the relationship between linguistic category and simple auditory discrimination, we employ both speech and non-speech stimuli. For all stimuli, we find a discrimination peak at the point in the continuum where a pitch fall changes to a pitch rise. This peak does not appear to be related to the category boundary for speech stimuli, as revealed in a labeling task. Discrimination was somewhat better for non-speech stimuli than speech.

## 1. Introduction

Speech perception involves the complex interplay of general purpose auditory perceptual mechanisms and speech-specific processing which together allow the recovery of both categorical information and a wealth of gradient, typically non-linguistic, information. The relationship between the two sets of mechanisms remains a central topic of investigation in phonetics, cognitive neuroscience and neurolinguistics.

In this paper, we examine the perception of a categorical distinction between certain question/statement pairs that differ physically only in the associated intonation contour. The pitch contour exhibits a high final rise for questions and a (less steep) fall for statements. This distinction is widely acknowledged to be a clear category distinction in English, German and Dutch, at least. It is of special interest to the investigation of the relationship between general purpose and speech-specific processing, not least because a single physical cue (to a first approximation) underlies the distinction. This contrasts with well-studied consonantal distinctions where a categorical distinction based on manner or place is signaled by a host of cues in parallel [7, 10, 12].

Several researchers have looked for the hallmarks of categorical perception of intonational contrasts [9, 8], and question/statement distinction in particular [14, 13]. Ladd and Morton [9] examined the distinction between ”normal” and ”emphatic” accent peaks in English. They found a well-formed S-shaped identification function on their labeling task, but did not find a clear peak in the discrimination function at the inferred category boundary. Furthermore, they observed an interesting and previously undocumented asymmetry in discrimination performance. Stimulus pairs in which the second member had the higher $F_0$ value (AB pairs) were discriminated with much more success than the reverse, BA, pairs.

Two studies have adopted the classical categorical perception approach to the question/statement distinction which we focus on here. In Schneider and Lintfert (2003) , listeners did standard identification and discrimination tasks where stimuli were derived from a recording of the sentence ”Steht alles im Kochbuch”. Identification results confirmed that distinct categories were involved, with individual category switches all lying within 2 steps along the continuum, which corresponded to a difference of less than 30 Hz at the stimulus endpoints. Discrimination results were less clear cut. There was a broad plateau in the middle of the continuum, with poorer discrimination for the more extreme stimuli. An AB/BA difference was also apparent, with worse discrimination for BA, as found also by Ladd and Morton. The link between inferred category boundary and discrimination performance was weak or non-existent, leading the authors to suggest that there might be a third, ‘hidden’, category between the falling statement and the sharply rising question. The co-existence of categorical and gradient phenomena as indexed by intonation has been highlighted by Gussenhoven [4].

In Remijsen and van Heuven (1999) , the same approach was taken with Dutch. Again, each subject exhibited a clear categorical response in the identification task, and again the discrimination functions did not support a standard CCP interpretation. In this case, along with the AB/BA asymmetry previously noted, there were two peaks in the discrimination function: one medially, corresponding roughly to the inferred category boundary, and one at the low end of the continuum, corresponding approximately to the point at which stimuli changed from a final fall to a final rise. Despite the author’s claim that their results provide a ”clear instance of categorical perception of an intonational contrast”, no satisfying account of either the low discrimination peak or the AB/BA asymmetry is provided.

These studies leave several questions unanswered. Although claims are made of ‘categorical perception’ for this particular contrast, the discrimination functions observed differ between studies, and in neither case is the category boundary clearly indicated by a discrimination peak. It has been suggested that the two categories of ‘Question’ and ‘Statement’ may not exhaust the interpretations possible within the continuum. Furthermore, it is evident that the interpretation of a given naturally occurring token as belonging to one category or the other will depend on cues other than pitch information alone [11].

The present study addresses these issue by using both speech and parallel non-speech stimuli in a discrimination task. If linguistic categories affect discrimination performance, as predicted by categorical perception accounts, then there ought to be a clear difference in the shape of the discrimination functions obtained for speech and non-speech stimuli. As in tra-

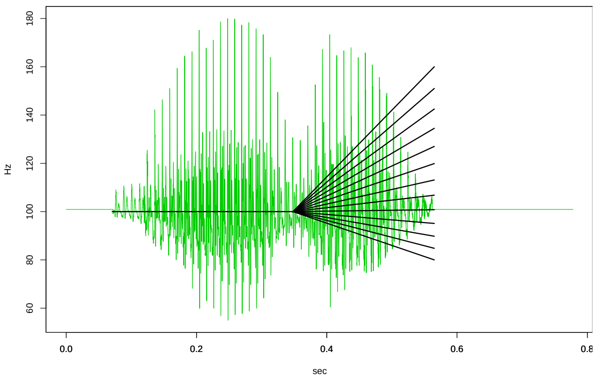

Figure 1: *Speech stimuli used with associated artificial pitch contours. Stimuli range from 0 (falling) to 12 (high rising). Stimulus 4 is essentially flat.*

ditional categorical perception studies, a labeling task can be used in conjunction with the discrimination task to evaluate the relationship between any discrimination peak and a category boundary. By using both speech and matched non-speech stimuli, we can also see if the AB/BA asymmetry previously reported for speech stimuli is specific to speech processing, or if its roots are to be sought in more general properties of auditory perception.

## 2. Methods

### 2.1. Stimuli

Four types of stimuli were employed, ranging from very speech-like to clearly non-speech. Initially, several repetitions of the first author repeating the isolated word ”Norway” with rising, falling and reasonably monotone intonation patterns were made[1]. These served to provide reference values for the endpoints of the stimulus continuum used, and one of the monotone recordings was selected as a model utterance for construction of all stimuli used.

For the speech stimuli, the model utterance was resynthesized with an intonation contour which was flat at 100 Hz over the first syllable and then descended or ascended linearly to a target value. The lowest target used was 80 Hz and the highest was an octave higher, at 160 Hz. Thirteen distinct points were used, with a one semi-tone difference between consecutive stimulus end points. The stimulus continuum is illustrated in Figure 1.

For the most speech-like of the non-speech stimuli, a single pitch pulse was excised from the original speech recording and reproduced many times over. This continuous voiced signal was amplitude modulated to match the original speech token, and the pitch contour resynthesized with the same values as the speech stimuli. This second stimulus set will be referred to as the voiced set.

A third set was made by generating a pulse train which was pitch synchronous with the speech stimuli used, and passing this through a series of linear filters representing five steady-state formants (Praat’s ‘To hum...’ command, [2]). The filtered sound was again amplitude shaped to match the speech original. This set will be called the hum stimuli.

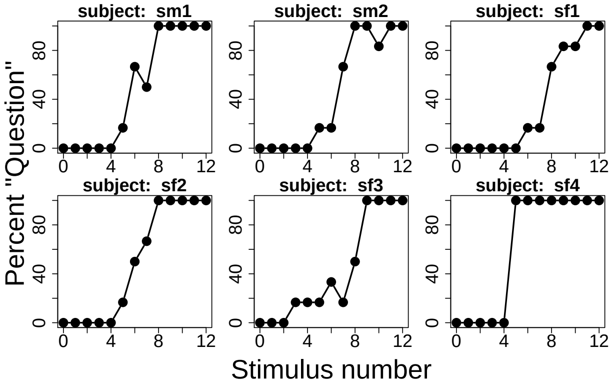

Figure 2: *Percentage of ‘Question’ responses as a function of stimulus number.*

Finally, a distinctly non-speech like set was generated in similar fashion, but without the formant-like shaping of the stimuli (Praat’s ‘To Sound (pulse train)’ command). These also had the same pitch pattern and amplitude contour as the speech originals. This final set will be called the buzz stimuli. Sample stimuli and discrimination pairs can be heard at [1].

### 2.2. Experimental Design

Six native English speakers participated (4f, 2m, ages 22–40). One female speaker was from North-West Canada. All other speakers were from the Republic of Ireland. Each subject participated in 4 one-hour trials which took place on distinct days. No subject reported any known speech or hearing deficit.

On each trial, two stimuli were played in succession and subjects performed a same/different forced choice task. Where the stimuli were different, they were adjacent stimuli within the 13 point continuum. Stimulus onsets were one second apart, and no repeat hearing was allowed. For each stimulus type, there were 24 possible ‘different’ trials, in which adjacent stimuli were presented, and these were randomly mixed with 26 ‘same’ trials, giving a basic per-stimulus type block of 50 trials. Sets of four blocks (one per stimulus type) were done consecutively, with a latin square ordering of sets among subjects. Four such sets could be completed in a single hour session, and subjects completed 4 sessions, giving a total of 3200 same/different discriminations per subject. Stimuli were played through Beyerdynamic DT 100 full cup headphones at a comfortable volume which was constant for all subjects in a quiet, but not sound-treated environment.

At the end of the fourth session, subjects completed an additional labeling trial in which they listened to each of the 13 speech stimuli 6 times in random order and labeled each as being either a question or a statement.

## 3. Results

In Figure 2, individual response functions are shown for the labeling task. The lower members of the stimulus continuum are all unambiguously labeled as ‘statements’, while the high members are labeled ‘questions’. The boundary between the two cat-

[1] Previous work had shown that a single word was sufficient to reliably capture a question/statement difference [3], and a short stimulus has the advantage of facilitating a large number of discrimination trials.

| sf1 | sf2 | sf3 | sf4 | sm1 | sm2 |
|---|---|---|---|---|---|
| 7.9 | 6.2 | 6.9 | 4.5 | 6.2 | 6.7 |

Table 1: *Estimate of category boundary obtained by probit analysis. Stimulus 4 is essentially flat.*

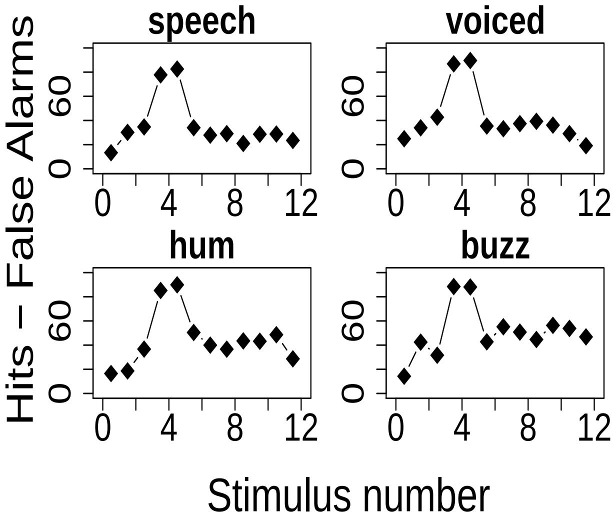

Figure 3: *Discrimination, indexed by number of hits minus number of false alarms.*

egories lies at approximately stimulus 7 for each subject, except subject sf4, who also has the clearest category boundary. Estimates of category boundaries obtained by probit analysis are provided in Table 1. The qualitiative boundary at which contours go from falling (stimulus 3), through flat (4) to rising (5) falls clearly within the ‘statement’ range for all subjects except sf4. It is worth noting that, although subjects were not selected on the basis of musical training, all had some musical training, and several played instruments regularly. Subject sf4, however, had considerably more formal musical training than any other subject (15 years) and was an active musician on a daily basis.

Figure 3 shows the discrimination performance as a function of stimulus number. The index of discriminability is given by the total number of hits minus the (interpolated) number of false alarms. The clearest feature of these data is the robust discrimination peak for stimulus pairs 3-4 and 4-5. Stimulus 4 is essentially flat, so these pairs are at the transition from a fall to a rise. This peak is clearly evident for all four stimulus types used, and may thus be taken to reflect acoustic discriminability, independent of linguistic categories. Analysis of individual subjects showed a consistent peak at the same location for all subjects, despite individual differences in response bias.

A finer breakdown of discrimination sensitivity is provided in Figure 4 in which hits for AB (higher stimulus is second) and BA are provided, along with false alarms from the ‘same’ trials. An ANOVA was performed on discrimination indices (hits minus interpolated false alarms) with factors of stimulus type, stimulus number and AB/BA order. This showed main effects of stimulus type [F(3,480)=7.1, $p<.01$], number [F(11,480)=22.5, $p<.01$] and order [F(1,480)=7.3, $p<.01$] with no significant interactions. Tukey HSD analysis revealed the non-speech ”buzz” stimuli to be significantly better discriminated than either the ”speech” or ”voiced” stimuli, and stimulus pairs 3-4 and 4-5 to be better discriminated than all other pairs. Examination of mean values showed the order AB (higher stimulus second) to be better discriminated than order BA.

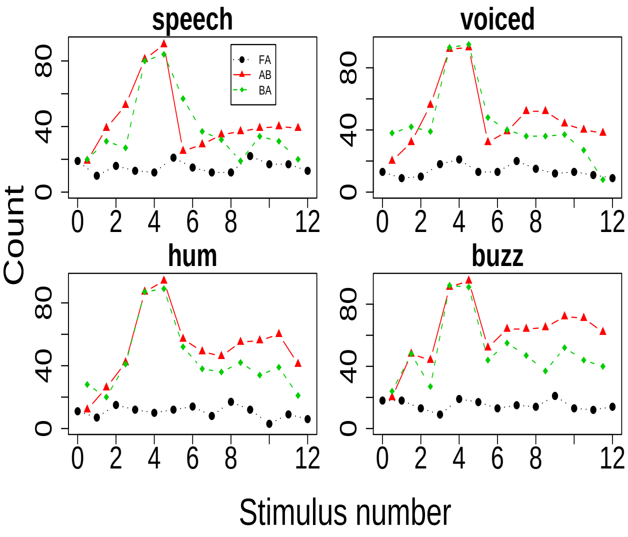

Figure 4: *Number of correctly discriminated ‘different’ trials for four types of stimulus. In AB/BA discrimination trials, B refers to the higher of the two stimuli. Also shown is the number of false alarms (FA).*

## 4. Discussion

Discrimination performance in this task does not appear to be influenced by the categorical nature of the question/statement distinction. Firstly, there is no simple relationship between the discrimination peak for the speech stimuli and the category boundary evidenced in the labeling task. The median category boundary as estimated by probit analysis was stimulus 6.7, while the peak was observed for stimulus pairs 3-4 and 4-5. Furthermore, the discrimination peak was a robust finding, clearly evident and invariant for all four stimulus types used, and for each subject individually. It is unsurprising that discrimination should be best at the transition from a falling contour to a rising contour. Cells which are specifically sensitive to either rises or falls are well documented throughout auditory cortex [15]. There is thus no motivation to relate this peak to any underlying linguistic categories. This seems to render untenable any attempt to describe the relation between pitch contour and linguistic category in terms of classical categorical perception [5, 11]. One subject, sf4, did exhibit qualitatively different behaviour from the others, in labeling all rising stimuli as ‘question’s and all flat or falling stimuli as ‘statement’s. The consistency of her responses, and the fact that she had by far the most musical training of all six subjects, suggests that she may have been responding directly to the fall/rise transition.

Not all of our findings are compatible with a simple auditory explanation, however. Discrimination was found to be better for non-speech stimuli than speech. We employed a range of four stimulus types, ordered to be progressively less speech-like, while maintaining the pitch contour and amplitude characteristics of the original recording. Despite these constants, the difference in source across the stimuli introduces numerous changes to the spectral/timbral characteristics of the tokens, making it impossible to judge whether the improved discrimination is due to the absence of linguistic categories, or to the spectral properties of the buzz stimuli.

Our findings are consonant with those of House [6], who reviewed several studies of level tones and contour tones. Findings summarized therein suggest that level tones (our stimulus number 4 and perhaps flanking stimuli 3 and 5) are processed differently from contour tones. Level tones are found to be perceived with great sensitivity, based on the psychophysical pitch discrimination abilities of humans, while contour tones appear to have a more complex target structure based on rate of change and the timing of change with respect to the associated segmental material.

Finally, we noted the asymmetrical discrimination performance which has been reported before [9, 13, 14], whereby pairs with the higher stimulus last are discriminated more readily than pairs with the higher stimulus first. This odd asymmetry is still unexplained. This is the first study we are aware of in which the asymmetry has been demonstrated for both speech and non-speech stimuli, and the absence of an interaction between stimulus type and presentation order suggests that the effect may be a general property of pitched stimuli, rather than speech-specific. A more extensive psychophysical investigation of the effect now seems warranted.

## 5. References

[1] http://cspeech.ucd.ie/ fred/intonation. Sample stimuli.

[2] Paul Boersma and David Weenink. Praat: doing phonetics by computer [computer program]. www.praat.org, 2005.

[3] C. Doherty, W. C. West, , L. C. Redi, D. Jr Gow, S. Shattuck-Hufnagel, and D. Caplan. The processing of question-intonation: an fMRI study. In *Proceedings of the 15th International Congress of the Phonetic Sciences*, pages 1647–1650, Barcelona, 2003.

[4] Carlos Gussenhoven. Discreteness and gradience in intonational contrasts. *Language and Speech*, 42(2–3):283–305, 1999.

[5] Stevan Harnad, editor. *Categorical Perception: The Groundwork of Cognition*. Cambridge University Press, 1987.

[6] David House. Perceptual thresholds and tonal categories. In *Proc. Fonetik*, pages 179–182, Umeå, 1997.

[7] Diane Kewley-Port, David Pisoni, and Michael Studdert-Kennedy. Perception of static and dynamic acoustic cues to place of articulation in initial stop consonants. *Journal of the Acoustical Society of America*, 73:1779–93, 1983.

[8] K. J. Kohler. Categorial pitch perception. In *Proceedings of the 11th ICPHS*, volume 5, pages 331–333, Tallinn, Estonia, 1987.

[9] D. Robert Ladd and Rachel Morton. The perception of intonational emphasis: continuous or categorical? *Journal of Phonetics*, 25:313–342, 1997.

[10] Leigh Lisker. *Rabid* vs. *rapid*: a catalogue of cues. *Haskins Laboratories Status Report on Speech Research*, 1985.

[11] Dominic W. Massaro. Categorical perception: important phenomenon or lasting myth. In *Proceedings of ICSLP*, pages 2275–2278, 1998.

[12] Louis C. W. Pols. Variation and interaction in speech. In Joseph Perkell and Dennis H. Klatt, editors, *Invariance and Variability in the Speech Processes*, chapter 7. Lawrence Erlbaum Associates, Hillsdale, NJ, 1986.

[13] Bert Remijsen and Vincent J. van Heuven. Gradient and categorical pitch dimensions in Dutch: diagnostic test. In *Proceedings of the 14th International Congress of Phonetic Sciences*, pages 1865–1868, San Francisco, 1999.

[14] Katrin Schneider and Britta Lintfert. Categorical perception of boundary tones in German. In *Proceedings of the 15th International Conference of the Phonetic Sciences*, pages 631–634, Barcelona, 2003.

[15] Shihab A. Shamma. Auditory cortex. In Michael A. Arbib, editor, *The handbook of brain theory and neural networks*, pages 122–127. MIT Press, 2003.